The Assembly and Structure of Self-assembling Peptides: Molecular to Supramolecular

Kyle Liam Morris

A thesis submitted for the degree of Doctor of Philosophy

Supervised by Professor Louise C Serpell

Department of Biochemistry

The University of Sussex

March 2012

Declaration

I hereby declare that this thesis has not been and will not be, submitted in whole or in part to another University for the award of any other degree.

The data and interpretation presented is the result of my own work unless done so in collaboration, as indicated in Chapters 6 – 8.

Signature:……

Kyle Liam Morris

Acknowledgements

Thank you Louise, unquestionably above all else, I owe you my unending gratitude. I have taken inspiration from you both professionally and personally. I cannot imagine anything that would teach me what you have. Your fluid but rigorous thought regarding science is a quality I admire and hope to have found myself. Balanced with your kindness and bright perspective on life, I am lucky to know you. You are, also, a fabulous singer.

Dave, thank you, I pay credit to your attitude to research: to not be thrown by unusual data, to make hypotheses and prove them wrong just as quickly, only in the pursuit of truth. Thank you to Lin for some very fun data collecting. Your immeasurable enthusiasm is wonderful and my opinion is forever changed on late night diffraction with take away pizza.

To my colleagues in Warwick and Birmingham, Matt and Tim, I am lucky to know your unlikely mix of brilliance, expertise, fun, and also our shared love of one cm thick bacon. Alison, our critical discussions of all things biophysical have been invaluable. Without these, this work would pale. A huge thanks to my colleagues in Norway who hosted me for a very memorable week, I sincerely hope to return. Pawel, Magnus and Florian, for your critique, interest and incredible help, I thank you – you are gentleman. In Denmark, Annette, our discussions of long-range order and modelling prompted so much. Joost, Fred and Kerensa, you helped to start all of this. At Sussex, Karen, thanks are not enough. You have left such an impression: who wouldn't with disco time, Mantis Shrimp and Watson & Crick moments. Ed, you have been an unassuming and extraordinary friend. Lisa, thank you for saving the port and for years of good office banter! Julian, you're such a delightful expert, thank you. My friends and colleagues – Daniel, Kathy, Martha, Tom, Patrick, Valia, Maris, Youssra, Aidan (wild thing), Nigel, Darren, Sharon, Sarah, Sunny, Ewan, JJ, Charlie, Wendy, Elodie, Nick & Tereza, Owen, Aline, Caroline, Kate, Fiona, Iain, Mark and Gareth – you all deserve so much gratitude, thank you. Simon, Aaron, Alastair, Martin, Daniela and Stephen, my best wishes and thanks.

To all those who made Brighton my home. Mary, Nic, Liz and Bod, thank you so much for all these years. We've come a long way together haven't we? Dibs and Dom, thank you for putting up with me, it is a real pleasure. Hetty, who never left this home. Your love and support shines through, thank you for this: our journey together, the treasure hunts, an abandoned pursuit of fun and of course, to be reminded that a good dance is the best medicine. Thank you to my family: Dad, Catherine, Gary, Claire, Barbie, Affa and Bonnie, where would I be? Though no words could do you justice, Dad you are a scholar and a saint.

Absolutely none of this would be possible without the patience and support of those that have loved, have shared themselves, their experience, or have embarked on the same journey as me. To these people, I am truly grateful.

So oft in theologic wars,
The disputants, I ween,
Rail on in utter ignorance
Of what each other mean,
And prate about an Elephant
Not one of them has seen!

The Blind Men and the Elephant
John Godfrey Saxe
(1816-1887)

University of Sussex

Kyle Liam Morris

A thesis submitted for the degree of Doctor of Philosophy

The Assembly and Structure of Self-assembling Peptides: Molecular to Supramolecular

SUMMARY

Self-assembling molecules are central to a plethora of processes found in nature, biotechnology and even disease. The importance of the non-covalent interaction of monomers to the formation of fibrillar assemblies is evident in the repeated use of this mechanism throughout nature, from essential cellular processes such as the formation of the cytoskeleton to the production of silk. Further, it has been recognised in the last two decades that a self-assembly mechanism, that is the formation of amyloid, underpins the pathology of protein misfolding diseases; it is therefore essential to dissect these mechanisms.

Despite recent technological and model system developments, self-assembling molecules remain challenging to investigate. Using combined structural and biophysical characterisations of penta- and hexa-peptide self-assembling model systems these investigations shed further light on the structure of amyloid-like fibrils. The elucidation of the structures of these fibrillar systems not only has implications for disease but also makes them well placed for consideration for biotechnological applications.

In reflecting upon how cross-ß structural architectures can be organised in the fibrillar state, a molecular and supramolecular model of fibrils formed by a fragment of α-synuclein is reported. The fibrils are found to consist of a novel and elaborate cross-ß architecture that leads to a helical supramolecular assembly spanning length scales previously unobserved for such a system.

Where self-assembly is a useful route to supramolecular structure formation, the use of low molecular weight gelator (LMWG) peptides to create fibrillar structures with defined material properties is also explored. The complex link between molecular structure, self-assembled architecture, fibril formation, fibril interaction and ultimately bulk material properties is described. It is found that the determinants of self-assembly are distinct from the determinants of gelation and so future LMWG design will have to consider both individually.

This work presents methodological advances in the characterisation of self-assembled structures. The investigations presented here have relevance for disease related processes but also to the technological use of these systems as materials. Finally, this work emphasises the beauty of the extravagant, yet elegant connection between molecular interaction and supramolecular self-assembly.

Table of Contents

Abbreviations

α-Syn	-	α-Synuclein
Aß	-	Amyloid-ß Peptide
AD	-	Alzheimer's Disease
ADP	-	Adenosine Diphosphate
AFM	-	Atomic Force Microscopy
ATP	-	Adenosine Triphosphate
APP	-	Amyloid Precursor Protein
CD	-	Circular Dichroism
CPL	-	Circularly Polarised Light
CSD	-	Cambridge Structural Database
DNA	-	Deoxyribose Nucleic Acid
ECM	-	Extra Cellular Matrix
EFF	-	Eclipsed face-to-face
FAP	-	Familial Amyloidotic Polyneuropathies
Fmoc	-	Fluorenylmethyloxycarbonyl Chloride
FFT	-	Fast Fourier Transform
FRET	-	Fluorescent Resonance Energy Transfer
FTIR	-	Fourier Transform Infrared Spectroscopy
GdL	-	Glucono-δ-Lactone
GFP	-	Green Fluorescent Protein
HT[V]	-	High Tension Voltage
IAPP	-	Islet Amyloid Polypeptide
LD	-	Linear Dichroism
LPL	-	Linearly Polarised Light
LMWG	-	Low-Molecular-Weight Gelator
MC	-	Merocyanine
NRMSD	-	Normalised Root-Mean-Squared Deviation
NMR	-	Nuclear Magnetic Resonance Spectroscopy
PEM	-	Photoelastic Modulator
PD	-	Parallel Displaced
PDB	-	Protein Data Bank
PMT	-	Photo-Multiplier Tube
PrP	-	Prion Protein
RNA	-	Ribose Nucleic Acid
SALSA	-	Simple Algorithm for Sliding Averages
SAF	-	Self-assembling Fibre
SAXS	-	Small Angle X-ray Scattering
SP	-	Spiropyran
TEF	-	T-shaped edge-to-face
TEM	-	Transmission Electron Microscopy
TTR	-	Transthyretin
WAXS	-	Wide-angle X-ray Scattering
XRFD	-	X-ray Fibre Diffraction
P-XRD	-	Powder X-ray Diffraction
SC-XRD	-	Single Crystal X-ray Diffraction

List of Figures

List of tables

List of equations & terms

1 Introduction I:

The self-assembly of biopolymers, aggregation and amyloid

1.1 Introduction

Proteins are crucial to the functioning of life. The central dogma describes the storing of genetic information in deoxyribonucleic acid (DNA), which is transcribed into ribonucleic acid (RNA) that is subsequently translated into polypeptide chains. This process is well understood and of course it is known that polypeptide chain synthesis is dependent on the genetic material coding for it. However, DNA and RNA are concurrently dependent on the proteins they encode for, for they are synthesised, maintained and regulated by the same proteins. Implied by the central dogma the function of DNA is predominantly comprised in its linear sequence of base pairing; but a polypeptide, although a great deal of information as to its function is encoded in its primary sequence, must undergo a complex folding process to gain its functionality. It is this folding process that is less well understood, and it remains one of the key fundamental problems of modern biochemistry.

1.2 Polypeptide folding

Anfinsen *et al.* demonstrated that a protein could be chemically denatured to result in a loss of its activity. Furthermore, in the absence of that denaturant the protein would spontaneously refold to regain its activity and function (Anfinsen et al., 1961). Two important conclusions can be realised from this observation. Firstly, that the structure of a protein is vitally important to its activity and function; and secondly, all of the information prescribing the three-dimensional structure of a protein is encoded by the primary amino acid sequence of a protein. Figure 1.1 shows the remarkable three-dimensional structure of green fluorescent protein (GFP) in an illustration of its unfolded and unstructured state and its experimentally determined natively folded and structured state. The three-dimensional structure of this polypeptide is a mixture of

the secondary structural elements, characterised as ß-sheets and α-helices. Seminal papers published by Pauling and Corey describe the adoption of these configurations by all polypeptide chains whereby the maximal use of available hydrogen bonding makes them the most stable conformations available to such molecules (Corey and Pauling, 1953, Pauling and Corey, 1953b).

This 26.9 kDa protein possesses an elegant arrangement of secondary structural elements where ß-strands form a ß-barrel conformation. The fluorescent properties of GFP are dependent on the 65-Ser-Tyr-Gly-67 motif that autocatalytically forms a 4-(*p*-hydroxybenzylidene)-imidazolidin-5-one structure (Cody et al., 1993). The formation of this novel fluorophore grants GFP its spectral properties but its fluorescent properties are abolished when it is unfolded. Inspection of the crystal structure reveals potential hydrogen bonding networks between the fluorophore and neighbouring polar residues, which would affect charge delocalisation and perhaps explain the varied spectral behaviours of GFP mutants (Yang et al., 1996).

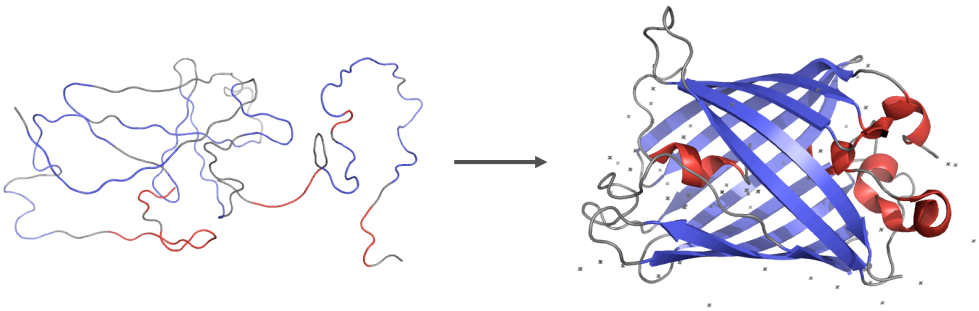


Figure 1.1: The folding and structure of a linear chain of amino acids in the three-dimensional structure of green fluorescent protein (1RRX), secondary structural elements are differentially labelled. Graphics generated in PyMol (DeLano, 2002).

Such a complicated arrangement must fold spontaneously and largely independently. Thus the native state should be more stable than the unfolded state. Protein folding has a high entropic cost in the ordering of the polypeptide chain and the ordering of water around a folded protein. In Equation 1.1, the Gibbs free energy (G) of a thermodynamic process is described. A decrease in entropy (ΔS) at constant temperature (T) will result in a positive contribution to the change in Gibbs free energy of protein folding (ΔG). Since for a thermodynamic process to be spontaneous the ΔG associated with it must be negative, where ΔS is negative for the folding process, so must the change in enthalpy (ΔH) be negative. In protein folding the large decrease in enthalpy is provided in the formation of intermolecular bonds (i.e. hydrogen bonds), which as they form lose energy to their surroundings.

$$\Delta G = \Delta H - T\Delta S \qquad \textit{Equation 1.1}$$

The interactions between the polypeptide chain must be more stabilising than the interactions with solvent, thus the native state and the conformations leading up to this will be more stable than their unfolded prerequisites. There are, however, many possible configurations of a polypeptide chain that may be spontaneously formed and be more stable than the unfolded state. If a polypeptide chain were to explore the entire conformational space leading up to the most stable native state then it would take longer than the time in the universe for a polypeptide chain to fold correctly (Levinthal, 1968). To rapidly fold on a realistic time scale (i.e. microseconds) polypeptide chains must selectively sample the conformational space on the pathway to adopting the native state. Considering the energetic preference of a polypeptide chain towards the native state and the necessary bias placed on a protein to fold quickly one can imagine the energy landscape associated with a folding polypeptide chain as a funnel, represented in Figure 1.2a.

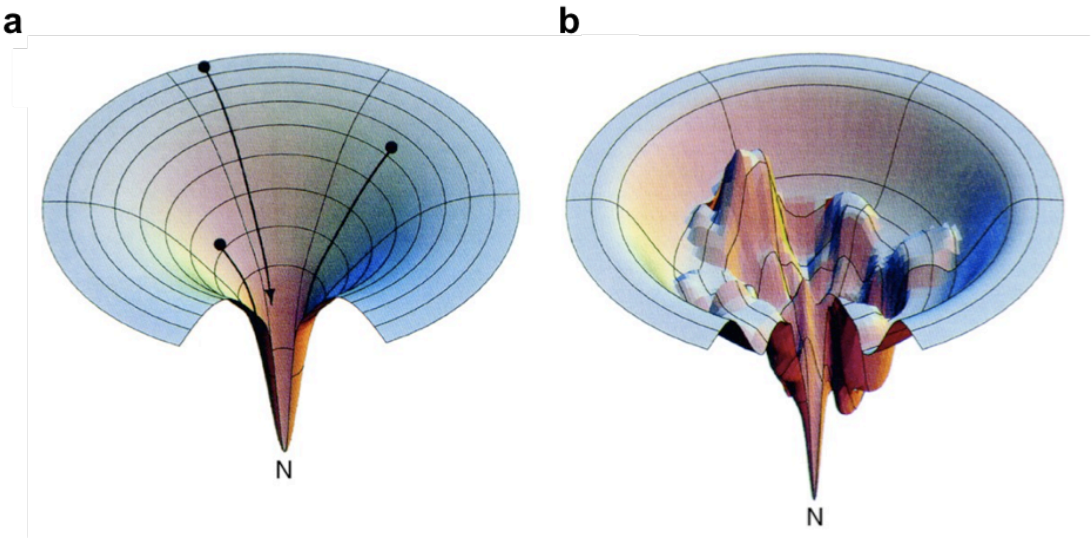


Figure 1.2: A representation of the energy folding landscape of a protein illustrating the bias towards the low energy natively folded state (a), and that although biased this landscape is rugged due to quasi-stable intermediates (b). Reproduced with permission from (Dill and Chan, 1997).

This perspective is useful but the new view of these protein folding energy funnels is better described in Figure 1.2b. Here the funnel will also cause a polypeptide chain to have a bias towards the natively folded state but the energy landscape towards this is actually quite complex where some folding intermediates, with respect to the unfolded state, will be marginally energetically stable. These partially stable intermediates may be necessary for the folding process where certain structural elements should form before others such that the final overall conformation is correct. Indeed, this has been demonstrated for the relatively simple protein hen egg lysozyme where particular secondary structural elements fold faster than others. This creates many possible parallel folding pathways with important intermediates but significant structural organisations allow these to converge onto the final required folded conformation (Radford et al., 1992).

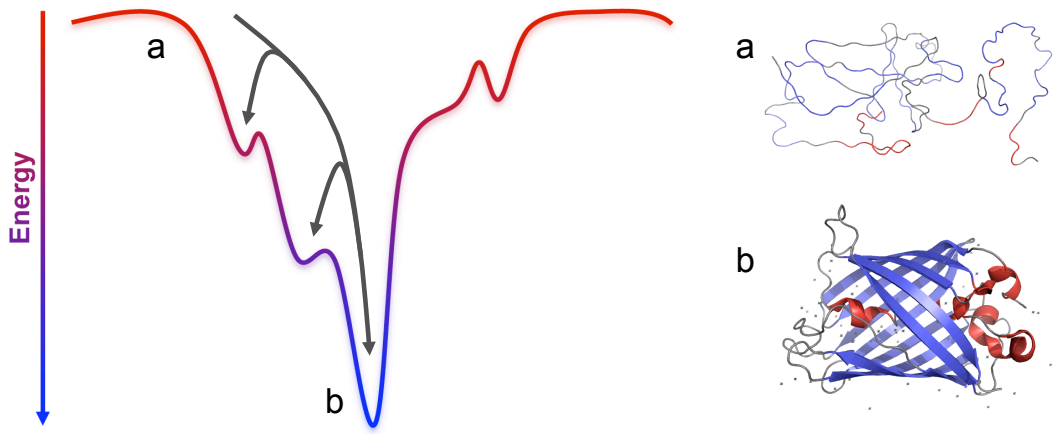


Figure 1.3: A cross section of a protein folding energy landscape. High energy unfolded states are represented in red and low energy folded native states in blue. A protein may follow a path through the energy landscape to reach the lowest energy natively folded state, either independently or assisted by chaperones. Inspection of the landscape reveals areas where energetically trapped intermediates may exist, these may be encountered on folding but changes in environmental factors such as temperature or pH can unfold a protein and allow it to access these intermediates. It has been suggested that the amyloid conformation represents a kinetically trapped species on this energy landscape (Hwang et al., 2004).

The example of GFP may be considered again, where Figure 1.3 shows a representational cross section of the energy landscape associated with the folding of a protein. The unfolded polypeptide chain (Figure 1.3a) moves through the energetic landscape towards the folded state (Figure 1.3b) possibly encountering the aforementioned intermediates. Additionally, changes in environment such as temperature or pH may allow a natively folded protein to access these intermediate states. In the case where energetically stable intermediates may be formed in an uncontrolled manner molecular chaperones have evolved to guide polypeptide chains to their native conformations (Hartl and Hayer-Hartl, 2009).

Figure 1.4 shows an illustration of the protein folding pathway *in vivo* and the possible fates of a translating polypeptide chain. The intermediately stable states may spontaneously form and be a necessary part of the folding pathway, or may represent kinetically trapped intermediates which molecular chaperones are designed to help avoid (Capaldi et al., 2002). If unwanted, the partially or misfolded protein may be assisted into the native state or targeted for degradation by the proteasome (Hartl and Hayer-Hartl, 2009).

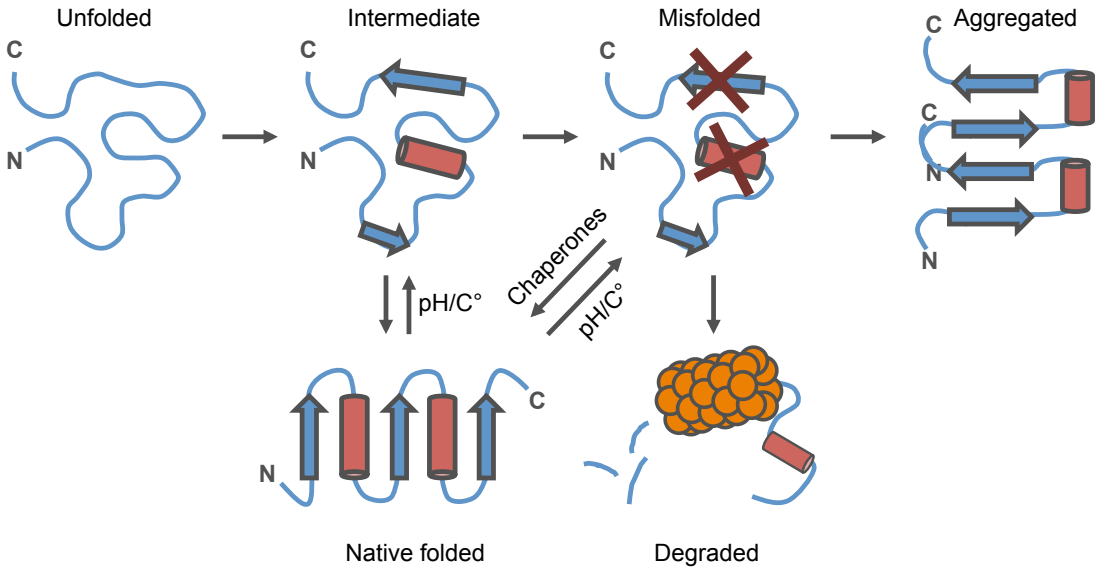


Figure 1.4: An illustration of the possible fates of a polypeptide chain post-translation in vivo. Polypeptide chains will fold to the natively folded state, possibly via intermediates but these intermediates may be energetically trapped and prevent proper folding in which case the protein would be targeted for degradation by the proteasome. Though misfolded proteins may accumulate and aggregate.

As stated, a protein may also access these misfolded conformations by a change in environment disrupting the native fold. If a protein accesses one of these conformations and it is not guided to the native state, efficiently refolded by chaperones or targeted for degradation by the proteasome, then it may lead to an off-pathway aggregative process (Hartl and Hayer-Hartl, 2009).

1.3 Polypeptide misfolding, aggregation and amyloid

Two classes of protein misfolding can be considered. In one case a polypeptide chain may be natively unstructured and aggregate; in the alternative, a protein may be natively folded, may unfold and subsequently misfold (Kelly, 1996). Native protein folding requires a delicate balance between the intermolecular forces stabilising the tertiary structure of the polypeptide backbone, interaction with solvent and other proteins. Globular proteins have a hydrophobic core and hydrophilic surface but if a protein is unable to adopt this general tertiary structure due to misfolding it may gain a propensity to self-associate in order to minimise hydrophobic –

water interactions. This is perhaps an oversimplified view of misfolding, but one that nonetheless illustrates how misfolding can result in protein aggregation. Protein aggregation might well represent an alternative-folding pathway of sorts, where conformations compatible with self-association are possibly an inherent property of the polypeptide backbone (Dobson, 2001). The conformation most commonly associated with self-association and self-assembly is the ß-strand. It has long been recognised that this structure could be adopted to form self-assembled ß-sheet rich assemblies called amyloid as described in the following sections.

1.4 Historical perspectives of amyloid

It was appreciated quite early on that a large number of globular polypeptides had access to an extended conformation stabilised by inter-amide backbone hydrogen bonding. This was originally explored using proteins such as ß-keratin, vegetable globulins and even denatured egg-white. The chosen technique used to study their structure was X-ray fibre diffraction (XRFD) (Astbury and Dickinson, 1935). These studies described groups of extended polypeptides arranged into what were described as “grids”, now known as ß-sheets. These models were found to be relevant to other proteins in their aggregated form but also proteins natively found in a fibrillar form such as silks. Models of *Bombyx mori* (Silkworm) silk have been proposed where extended polypeptide chains are arranged with their long axes running parallel to the fibre axis of the silk (Marsh et al., 1955). It was work on *Crysopa flava* (Green lace-wing fly) egg-stalk silk that introduced the cross-ß concept (Geddes et al., 1968). The cross-ß architecture is comprised of ß-strands arranged with their long axes perpendicular to the fibre axis where the regular and repeating hydrogen bonding distance along the fibre axis stabilising the ß-strands into ß-sheets is 4.7 Å. ß-sheets may then laterally pack perpendicular to the fibre axis separated by a distance of 9.5 Å.

Prior to this, Cohen and Calkins made observations by light microscopy studies on sections of tissue taken from systemic amyloidosis patients finding fibrillar deposits known as amyloid

(Cohen and Calkins, 1959). Based on histological staining they noted a distinction between the *in vivo* native fibrils and were able to identify the fibrillar amyloid deposits to make some limited assessment of their morphology. Soon after, further morphological information was ascertained after purification, isolation and negative stain transmission electron microscopy (TEM) of the previously reported amyloid fibrils. With the fibrils in isolation these investigations were able to appreciate that amyloid fibrils were long, unbranched and have a hierarchical structure whereby mature fibrils are made up of tightly packed, laterally associated protofilaments (Shiraham and Cohen, 1965).

Eanes and Glenner then showed amyloid filaments purified from the spleen and liver from patients with systemic amyloidosis exhibited the XRFD pattern typically associated with the cross-ß architecture (Eanes and Glenner, 1968). The authors noted the similarity to previous studies that had investigated this conformation. Although the cross-ß architecture was thought to be widely accessible to polypeptides if treated in the correct harsh manner Eanes and Glenner noted that their methods of extraction were relatively mild. Extracts of Bence Jones immunoglobulins from patients with no known amyloidosis could be treated in such a way that by TEM they appeared fibrillar and exhibited a typical cross-ß diffraction pattern by XRFD (Glenner et al., 1971). However due to the mild treatment by Eanes and Glenner, they made the first suggestion that a polypeptide was accessing the cross-ß amyloid structural space independently and to the detriment of the host organism.

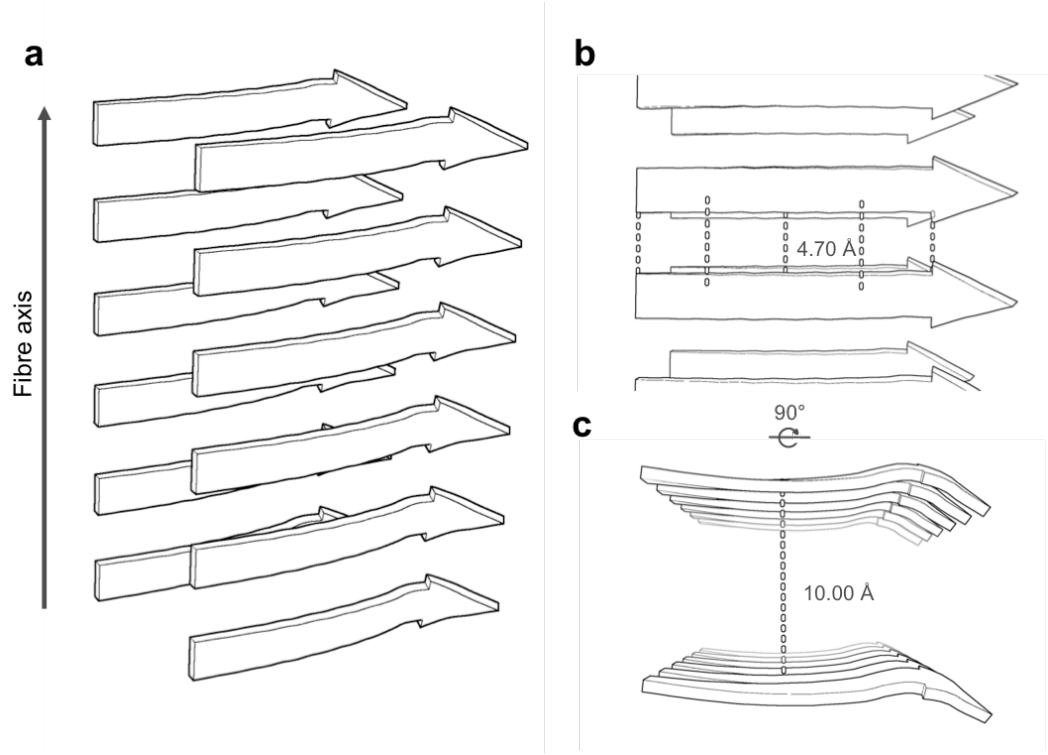


Figure 1.5: A representation of the Kirschner cross-ß model (a), showing the ß-strand separation of ~4.7 Å (b) and ß-sheet separation of ~10 Å (c). Graphics generated in PyMol (DeLano, 2002).

After these initial observations and the ability to extract amyloid from tissues, the relationship between the long un-branched fibrils observed by TEM and the cross-ß architecture determined by XRFD was clear. Furthermore, the connection between the cross-ß architecture, pathological fibrillar deposits and thus their implication in disease has resulted in a large field investigating their pathological role. Continued intense research into the cross-ß conformation in amyloid proteins resulted in the collection of an XRFD pattern from *ex vivo* fibrils associated with Alzheimer's disease, Amyloid-ß (Aß). Here again, the characteristic spacings associated with an arrangement of stacked ß-strands were seen with a major meridional reflection at 4.76 Å and a diffuse major equatorial reflection 10.6 Å (Kirschner et al., 1986) as shown in Figure 1.5. Shortly after, synthetic fragments of the full-length Aß protein were synthesised and found to assemble in the same manner. The synthesis of the Aß fragments and their corresponding cross-ß XRFD patterns definitively placed this protein and its aggregation as a central cause of

Alzheimer's disease, but also represents an important step that is the beginning of the use of model systems to study amyloid structure by XRFD (Kirschner et al., 1987).

1.5 Amyloid and disease

The spontaneous self-assembly and aggregation of polypeptides into higher order structures, called amyloid, is associated with a number of (neuro)degenerative diseases. Parkinson's disease is associated with the aggregation of α-synuclein, Huntington's disease with expanded polyglutamine caused by genetic CAG repeats, Alzheimer's disease (AD) by an aggregation of Aß peptide, diabetes with accumulation of islet amyloid polypeptide and a number of transmissible spongiform encephalopathy prion diseases. Whilst these diseases may be characterised by the presence of their corresponding proteins, the precise correlation between aggregation and disease pathology is not entirely clear and still the subject of intense investigation.

In particular, AD is characterised by observable pathological features including diffuse and senile plaques in the brain tissue, amyloid is also present in the cerebrospinal fluid and neurofibrillary tangles in cell bodies comprising of hyperphosphorylated tau protein (Selkoe, 1990). Sometimes referred to as neuritic plaques, senile plaques are spherical and have a diameter of 20 – 50 μm with a dense core abundant in Aß (Morgan et al., 2004). Identification of the rich Aß core has led to a focus of research around this peptide itself in the hopes of developing understanding of the pathology of the disease. Aß is a ~4kDa protein with a common conserved core sequence but varying a variable C termini. It is derived from a membrane bound protein, amyloid precursor protein (APP). APP is a large naturally occurring protein and its native biological function is not fully understood. APP can be cleaved by a group of enzymes called the aspartyl proteases, differential cleavages result in two main forms of Aß being produced: Aß(1-40) and Aß(1-42), the latter of which is overall less abundant (Irvine et al., 2008). Investigations have shown evidence that Aß(1-42) has a greater propensity to

aggregate than Aß(1-40), implicating the ratios of Aß(1-42)/(1-40) as being important in AD patients but ultimately in either form the spotlight was placed on Aß as the principle molecular cause of AD. Hardy and Higgins proposed the amyloid cascade hypothesis as a molecular explanation for the development of AD (Hardy and Higgins, 1992). The cascade hypothesis later received criticism where it originally described Aß fibrils as the central cause of AD, but it has since been refined to place soluble oligomeric Aß as the primary pathogenic cause of AD (Hardy, 2002). However the presence of Aß alone cannot be causative of neurodegeneration because healthy individuals produce Aß in the pico- to nanomolar concentrations, rather it is the ordered self-association of Aß molecules that causes AD pathology.

Studies have demonstrated that synthetic Aß peptide that is allowed to assemble *in vitro* is directly toxic to neurons (Deshpande et al., 2006). Varied hypotheses from a range of studies have been proposed for the exact mechanism of Aß toxicity, ranging from disruption of cellular ionic homeostasis (Bhatia et al., 2000, Mattson et al., 1992), oxidative stress (Yatin et al., 1999), membrane permeation via specific ion channel formation (Durell et al., 1994), membrane permeabilisation through fibril fragmentation (Xue et al., 2009) and more general interactions (Sokolov et al., 2006, Williams and Serpell, 2011), to interference in cell signalling (Resende et al., 2008). These studies each subtly implicate different aggregation states in the toxic mechanism, but at the same time are not in all cases able to discriminate between the principle toxic species. Thus it is beneficial to better understand the aggregation pathways of such proteins.

1.6 Amyloid aggregation pathway

Like a protein folding pathway the aberrant aggregation of polypeptides has classifiable intermediates. As shown in Figure 1.6 these occur by a nucleated polymerisation mechanism (Ferrone, 1999). The lag phase corresponds to a period of low fibril elongation where nucleation events create extension-competent structures, which may become fibrils. Primary nucleation events may occur all the time monomer is available but fibril fragmentation is now also recognised as a secondary mechanism for initiating fibril growth (Xue et al., 2010).

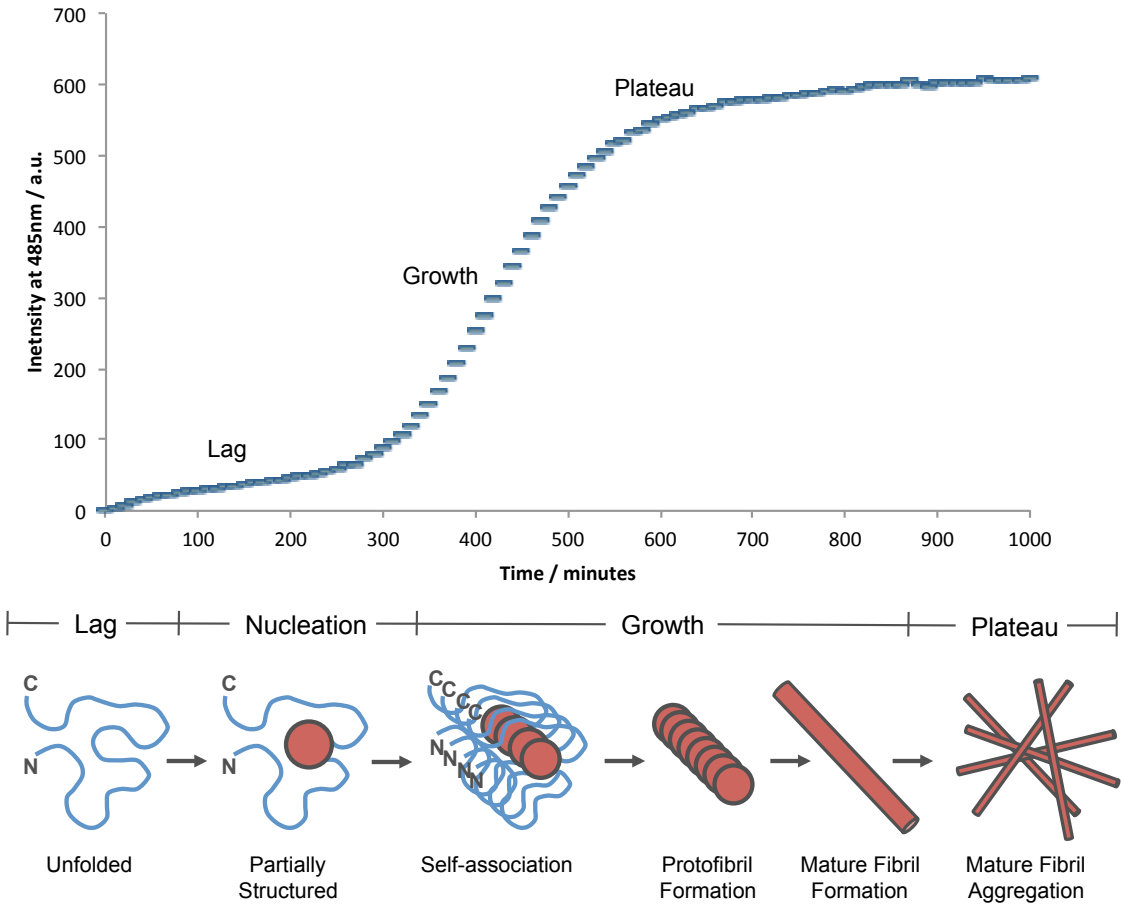


Figure 1.6: An illustration of the amyloid assembly pathway. Thioflavin-T fluorescence assays can be used as a fluorescent reporter for amyloid formation revealing the nucleated growth kinetics of the assembly mechanism. As such, the lag, growth and plateau phases can be correlated with particular structural species.

As described, aggregates of polypeptides can be loosely grouped into fibrillar and non-fibrillar states. The non-fibrillar states may be intermediates on pathway to fibrillar states or amorphous aggregates. Amorphous aggregates cannot strictly be described as amyloid though as this is defined by a very specific misfolded conformation. The amyloid fibrillar state is hierarchical and best described as filamentous. The descriptions of these intermediates are sometimes confusing in the literature and so it is useful to summarise them here as in (Murphy, 2007):

1.6.1 Non-fibrillar:

Monomers may be structured or unstructured where some amyloid precursors are natively folded polypeptides such as Transthyretin (TTR) or Prion Protein (PrP) whereas others are natively unfolded such as Aß (Kelly, 1996). Natively folded precursors undergo a conformational change to become amyloid. They may have either α-helical secondary structure content and be converted to ß-sheet or have native ß-sheet content and undergo rearrangement into an amyloid architecture. Unstructured precursors appear to have a semi-stable state that is dynamically dependent on the environment. There have been reports of monomeric Aß containing some ß-strand structure but contrasting results have shown the adoption of two parallel α-helices when observed in an environment said to mimic the conditions at the plasma membrane (Shao et al., 1999).

Oligomers are globular aggregates that like monomers may have transient and poorly definable structure. They appear in many *in vitro* preparations of Aß as spherical structures with diameters of between 5 – 15 nm. The oligomers likely represent the necessary nucleus that must form from which conversion to fibrillar material can proceed (Lee et al., 2011), although the nucleation of fibril growth can be through other mechanisms including fibril fragmentation (Xue et al., 2008).

Non-fibrillar aggregates are considered to be the amorphous mass that occurs alongside other structural forms of amyloid. They lack any obvious structure as compared with other amyloidogenic species although they may possess some ß-sheets and often predominate before solutions of amyloidogenic peptide stabilise into more structured forms.

1.6.2 Fibrillar:

Protofibrils are considered to be the early intermediates in the amyloid formation pathway. The term is used broadly in the literature with structural descriptions being varied. Protofibrils can grow up to 11 nm in diameter and are generally <200 nm in length (Murphy, 2007). They are not as structurally defined as more mature fibrils but exhibit a repeating monomer structure that gives them greater structural definition than non-fibrillar species (Harper et al., 1999).

Mature fibrils are fully mature amyloid structures of 7 – 20 nm in diameter and an indeterminate length where internal secondary structure is revealed to be infinitely propagating ß-sheet (Makin and Serpell, 2005a). Mature fibrils are linear and are likely composed of a number of *protofilaments*.

Protofilaments are the narrower filamentous subunits that laterally associate to form a mature amyloid fibril. They have a well-defined ß-sheet-rich core as described by many cryo-EM studies of amyloid fibrils (Jimenez et al., 1999, Jimenez et al., 2002, Serpell and Smith, 2000, Tattum et al., 2006).

1.7 Amyloid promiscuity

Amyloid fibrils represent an incredible structural architecture. The structural promiscuity of this architecture is evident by the sheer number of proteins and peptides that can access this three-dimensional conformation. A subject of intense multidisciplinary study, the key to understanding the nature of these proteins and peptides most likely lies in more than one

particular viewpoint or piece of understanding. However one such viewpoint has aimed to identify structure-function relationships to gain insight into these systems. Indeed, from a structural point of view, a great deal of information has been collected about amyloids. Due to the number of precursors that are known to be amyloidogenic, whilst there remains a great deal to understand, a rich description of the amyloid structural space has been developed. Table 1.1 shows the diversity of sequences that have been found to form amyloid and amyloid-like structures, as well as the broad range of protein and peptide precursors. Amyloid structures may arise from proteins that are natively unfolded, unfold and then misfold or have an inherent amyloid like structure. Amyloid fibrils are insoluble by nature, generally paracrystalline (having only long-range crystalline order in one dimension) and also highly heterogeneous. This has unfortunately meant that the majority of studies attempting to determine precise structural atomic co-ordinates of the amyloid state have been problematic. In contrast, studies that have attempted to understand the monomeric prefibrillar aggregates have had access to the standard suite of structural biological tools. In this way, most of the native structures of amyloidogenic proteins and peptides have been solved using solution nuclear magnetic resonance (NMR) or single crystal X-ray diffraction (SC-XRD). In many cases it is important to understand the prefibrillar monomeric structure of an amyloidogenic protein to illuminate the mechanisms that lead to fibril formation with a view to control, intervene or design these very mechanisms.

			Precursor Structure			Amyloid Structure			
Architectur	**Precursor**	**UniProt**	**Segment**	**Technique**	**PDB**	**Segment**	**Technique**	**PDB**	**Reference**
Cross-ß	TTR	P02766	2-147	SC-XRD	1F41	105-115	ssNMR	1RVS	(Gustavsson et al., 1991; Jaroniec et al., 2004)
	CA150	O14776.2	-	-	-	428-464	ssNMR	2NNT	(Ferguson et al., 2006; Neri et al., 2001)
	Poly-Q	P42858	1-64	SC-XRD	3IOT	$(D_2Q_{15}K_2)$	XRFD	-	(Macdonald et al., 1993; Sikorski and Atkins, 2005)
	Aß A4	P05067	(1-40)	NMR	1AML	(1-40)	ssNMR	-	(Petkova et al., 2002; Sticht et al., 1995)
	Aß A4	P05067	(1-40)	NMR	1AML	$(1\text{-}40)_3$	ssNMR/MPL	-	(Paravastu et al., 2008; Sticht et al., 1995)
	Aß A4	P05067	(1-40)	NMR	1AML	(11-25)	XRFD	-	(Sikorski et al., 2003; Sticht et al., 1995)
	Aß A4	P05067	(1-40)	NMR	1AML	(37-42)	SC-XRD	2ONV	(Sawaya et al., 2007; Sticht et al., 1995)
	Aß A4	P05067	(1-42)	NMR	1IYT	(1-42)	H/D NMR	2BEG	(Luhrs et al., 2005; Picone et al., 2002)
	ß2M	P61769	21-119	SC-XRD	1A1M	20-41	ssNMR	2E8D	(Iwata et al., 2006; Smith et al., 1996)
	IAPP	P10997	-	-	-	34-70	ssNMR	-	(Tycko et al., 2007)
	IAPP	P10997	-	-	-	(20-29)	ssNMR	-	(Madine et al., 2008)
	IAPP	P10997	-	-	-	(19-29)	XRFD	3DGJ	(Eisenberg et al., 2008)
	IAPP	P10997	-	-	-	(30-35)	XRFD	3DG1	(Eisenberg et al., 2008)
	Insulin	P01308	90-110	NMR	1AI0	35-41	SC-XRD	3HYD	(Chang et al., 1997; Ivanova et al., 2009)
	Sup35	P05453	-	-	-	7-13	SC-XRD	1YJO-P	(Wiltzius et al., 2009)
	Het-S	Q03689	-	-	-	218-289	ssNMR	2RNM	(Wasmer et al., 2008)
	ccßMet	N/A	N/A	N/A	N/A	N/A	TEM/XRFD/ssNMR/CP	-	(Steinmetz et al., 2008)
	$KF_2EA_3K_2F_2E$	N/A	N/A	N/A	N/A	N/A	ED/XRFD	2BFI	(Makin et al., 2005)
Helix / Solenoid	Aß A4	P05067	(1-40)	NMR	1AML	(16-22)	XRFD	-	(Mehta et al., 2008; Sticht et al., 1995)
	TTR	P02766	2-147	SC-XRD	1F41	N/A	XRFD	-	(Blake and Serpell, 1996; Gustavsson et al., 1991)
	Poly-Q	P42858	1-64	SC-XRD	3IOT	$(D_2Q_{15}K_2)$	XRFD	-	(Macdonald et al., 1993; Perutz et al., 2002)
	PrP	P23907	182-233	SC-XRD	1UW3	89-227	TEM/CP	-	(Bayley et al., 2004; Govaerts et al., 2004)
Native-like	TTR	P02766	2-147	SC-XRD	1F41	1-147	SC-XRD	1G1O	(Gustavsson et al., 1991; Sauer-Eriksson et al., 2000)
	P22-tail	P12528	119-666	SC-XRD	1TSP	119-666	SC-XRD	1TSP	(Steinbacher et al., 1994)
	Ure2p	P23202	95-354	SC-XRD	1G6W	1-354	TEM	-	(Melki et al., 2001; Steven et al., 2004)
	ß2M	P61769	21-119	SC-XRD	1A1M	21-116	SC-XRD	1LDS	(Smith et al., 1996; Trinh et al., 2002)
	ß2M	P61769	21-119	SC-XRD	1A1M	$[21\text{-}119]_6$	SC-XRD	3CIQ	(Miranker et al., 2008; Smith et al., 1996)

Table 1.1: A cross-section of the diversity of amyloid precursors and structures. Amyloid segments in some cases are referred to by their fragment names, indicated by parenthesis. NMR – solution state NMR, ssNMR – solid state NMR, H/D NMR – hydrogen-deuterium exchange NMR, MPL – Mass per length, ED – Electron diffraction, CP – Computational, XRFD – X-ray fibre diffraction, SC-XRD – Single crystal diffraction, TEM – Transmission electron microscopy, '-' – Not available or deposited at this time.

The much broader range of techniques used to characterise the amyloid state reflects the multi-technique approach required to characterise these challenging proteins. Only in the last ten years have high-resolution ssNMR and SC-XRD descriptions of the amyloid architecture become available and these have confirmed and extended our understanding of the amyloid conformation. Yet these represent only a small portion of the amyloid structural space and do not always describe amyloid formed by full-length proteins, for example: those associated with disease. X-ray fibre diffraction and transmission electron microscopy have proved to be excellent in complementarity to other structural techniques. The following sections describe recent advancements in our understanding of the amyloid conformation from XRFD, TEM, ssNMR and SC-XRD.

1.8 Amyloid structure

1.8.1 Fibre diffraction & TEM

As described in Section 1.4, fibre diffraction in many ways established our fundamental understanding of the structure of amyloid fibrils. Recently, fibre diffraction has made contributions individually, and in combination with other techniques, to insights into the structure of amyloid-like fibrils ranging from small proteins to short-peptide model systems. Whilst full-length amyloid proteins are difficult to study structurally, many structures for amyloid fibrils have recently been determined using fragments or short peptides extracted from their full-length amyloid sequence. As shown in Figure 1.7, decreasing the size of the system being studied generally increases the amount of structural information obtainable in XRFD data. These studies often take advantage of or identify the minimal sequence regions required for amyloid formation. Figure 1.8 shows a sample of the model structures of short-amyloidogenic peptides recently elucidated for amyloid where XRFD was the principle technique utilised. These all show the cross-ß architecture but with differing lateral packing arrangements, ranging from variable short-range interactions (side chain association) to long-range interactions (helical arrangements).

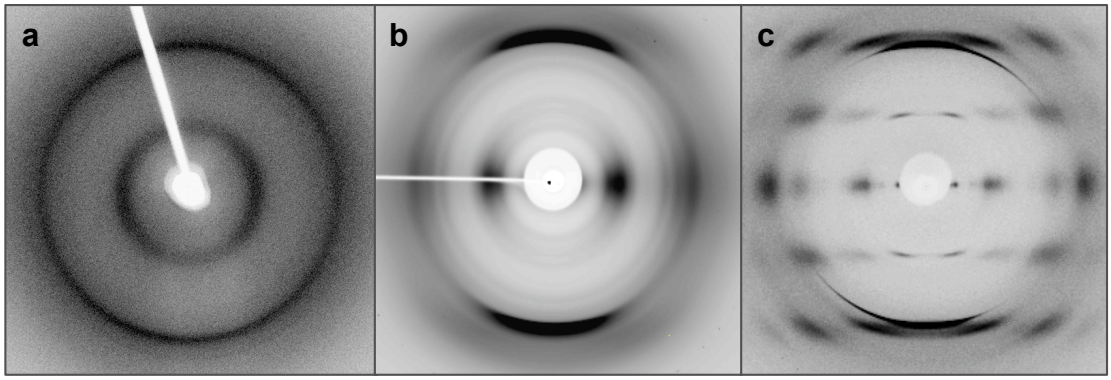


Figure 1.7: Fibre diffraction typically produced from alignments of amyloid fibres of Aß(1-40), Aß(1-42) and Aß(11-25). Full-length Aß(1-40) is often poorly aligned/disordered resulting in few diffraction signals, those that are present tend to be radially averaged and/or weak and diffuse (a), showing Aß(1-42) it is clear that in some exceptional cases better quality alignments may be obtained (b) (Sikorski et al., 2003) but shorter fragments, in this case Aß(11-25) routinely give high quality fibre diffraction patterns (Sikorski et al., 2003). All fibre axes are vertical. b and c reproduced with permission.

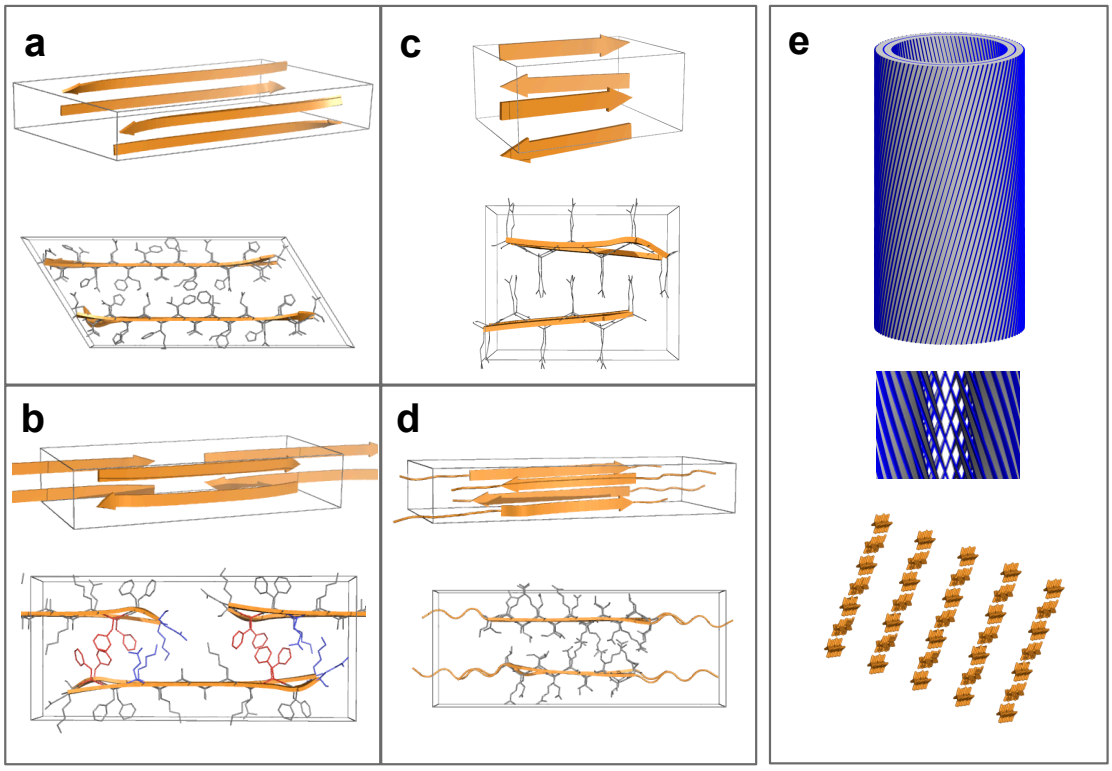


Figure 1.8: A range of amyloid models solved principally using fibre diffraction. a-Aß(11-25) (Sikorski et al., 2003), b-KFFEAAAKKFFE (Makin et al., 2005), c-PolyQ (Sikorski and Atkins, 2005), d-ccbMet (Steinmetz et al., 2008) and e-Aß(16-22) in nanotubular form. Panel e provided by Mehta reproduced from (Mehta et al., 2008). Figures reproduced from (Morris and Serpell, in press). Graphics generated in PyMol (DeLano, 2002).

Aß(11-25) is the central region of the full-length Aß protein associated with the development of Alzheimer's disease. XRFD revealed that adjacent ß-sheets slip by one residue relative to one another to allow for tighter interdigitation of side chains (Figure 1.8a) (Sikorski et al., 2003), an observation supported by more recent crystallographic studies of amyloid-like peptides (see Section 1.8.3) (Sawaya et al., 2007). Based on TEM micrographs and the proposed model for Aß(11-25), Sikorski *et al*. proposed that full-length Aß(1-40) consists of a ß-hairpin, ssNMR measurements made at the same time support this model (see Section 1.8.2) (Petkova et al., 2002).

The importance of side chain interactions has been emphasised in a number of studies (Marshall et al., 2010), in particular the role of aromatics in amyloid fibril formation (Gazit, 2002). Models that were elucidated using fibre diffraction have further highlighted these concepts. The designer 12-mer KFFEAAAKKFFE sequence contains diphenylalanine motifs notably found in the protein Aß and thought to play a key role in the assembly process (Gazit, 2002). X-ray fibre diffraction models of this peptide reveal aromatic stacking of phenyalanines but additionally electrostatic interactions are observed between adjacent Glu-Lys pairs (Figure 1.8b) (Makin et al., 2005). Mutagenesis studies using this system have consolidated the absolute importance of the phenylalanines in driving assembly via principally hydrophobic interactions but some directionality due to aromaticity. Additionally, mutagenesis of the charged residues Glu and Lys and investigations in different environments reveals that the charge-charge interactions modulate fibril packing and morphology whilst allowing the amyloid system to structurally adapt to a range of environments (Marshall et al., 2011). These studies illustrate the growing trend to remove partial sequences from full-length proteins or design new amyloidogenic sequences based on well-known amyloidogenic motifs to produce model systems.

X-ray fibre diffraction studies on the poly-glutamine expanded model protein $D_2Q_{15}K_2$ have modelled the core region of these amyloid fibrils to a 6-residue cross-ß peptide arrangement (Sikorski and Atkins, 2005). The structure shown in Figure 1.8c produces simulated fibre

diffraction patterns that match with experimentally collected data very well; and this perhaps implies the core region is responsible for the major features of any amyloid XRFD pattern. A designed cross-ß core peptide called ccß-Met was also solved using in part XRFD (Figure 1.8d); here the close match of the model to experimental data is a testament to the power of XRFD alone and in combination with other techniques (Steinmetz et al., 2008). Certainly it is well established from XRFD and other techniques that the cross-ß architecture forms the core of all currently studied amyloid polypeptides (Jahn et al., 2010).

As described, X-ray fibre diffraction can give structural information on the short-length scale as in these studies but the resolution range allows for long-range order to also be investigated. The supramolecular structure of amyloid fibrils has not only significance for understanding their stability but for nanotechnological applications (Baldwin et al., 2006, Banwell et al., 2009, Papapostolou et al., 2007) and for gaining a fundamental understanding of the self-assembly process. Using fibre diffraction, helical models representing fibril supramolecular structures have been proposed. Recent studies using the fragment Aß(16-22) have, using X-ray fibre diffraction, proposed underlying structural arrangements that lead to long-range morphological differences. This peptide has the unique property of forming either monodisperse fibrils or nanotubes under specific assembly conditions. As such, it was possible to obtain XRFD patterns from each morphologically distinct species. The two models proposed show ß-strands arranged approximately perpendicular to the fibre axis but in the case of the nanotubular morphology (Figure 1.8e) the hydrogen bonding between adjacent ß-strands is tilted 23° from the fibre axis introducing a helical rise per turn that extends into a nanotubular conformation (Mehta et al., 2008).

These represent a cross-section of the recent advances in our structural understanding of the amyloid architecture where fibre diffraction has been the principle technique used but a number of other reports where XRFD has been used in combination with other structural techniques have been reported including; Islet amyloid polypeptide (IAPP) (1-37) fibrils using electron

diffraction and XRFD (Makin and Serpell, 2004), IAPP(20-29) fibrils using a combination of ssNMR and XRFD (Madine et al., 2008), PrP using a combination of XRFD, computation and microscopy (Govaerts et al., 2004, Wille et al., 2009) and β_2-m(20-41) fibril using ssNMR, atomic force microscopy and XRFD (Iwata et al., 2006).

X-ray fibre diffraction gives insight into the interactions that stabilise these amyloid fibrils, but to truly understand the context of these interactions in the fibrillar form, the long-range order of these systems may be investigated, as such, various electron microscopic techniques have contributed significantly to our understanding. Figure 1.9 shows typical transmission electron micrographs collected for two different amyloid systems; insulin (Figure 1.9a) and Aß (Figure 1.9b) where the sample is negatively stained. Clearly the samples are both morphologically polydisperse ranging from thin and thick filaments with subtle indications of helical repeats, to obviously helical fibrils sometimes appearing as twisted ribbons, and in other instances as tightly wound ropes (morphologies described as in (Jimenez et al., 2002)). By comparison both have a similar polydispersity with mostly morphologically similar species. Certainly it would be challenging to distinguish the two samples from one another without foreknowledge. It is a fascinating point to consider that two entirely distinct proteins can adopt the same molecular architecture but because of this perhaps unsurprising that their supramolecular structure and macromolecular appearance be similar. However it raises questions over how exactly the cross-ß architecture is arranged within these fibrils and what insight can be gained from understanding this?

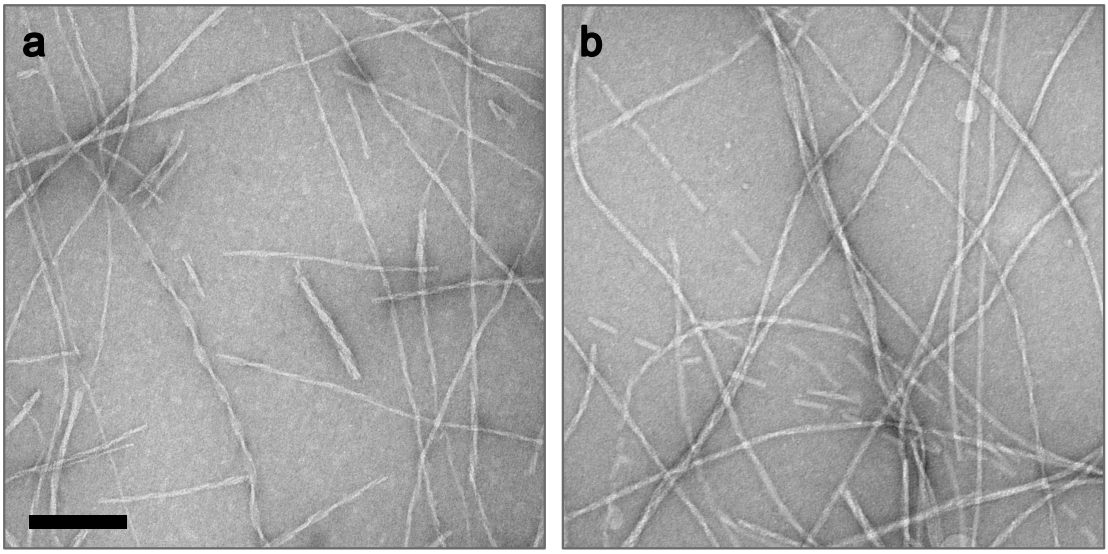


Figure 1.9: Comparison of transmission electron micrographs from two distinct amyloid systems, insulin (a) and Aß(1-40) (b), reveals varied but indistinguishable morphology. Scale bar represents 200 nm. Reproduced from (Morris and Serpell, in press).

An early example investigating the protofilamentous structure of amyloid fibrils made cross sections of *ex vivo* TTR(Met30) fibrils from patients with familial amyloidotic polyneuropathies (FAP) and used thin film sectioning techniques to visualise their cross sectional structure (Serpell et al., 1995). This technique was subsequently applied to a range of amyloid systems revealing the protofilament constituents of the mature amyloid fibril to be variable with respect to the packing arrangements adopted (Serpell et al., 2000b). The broad range of systems studied indicated that that the structural similarity found in all amyloid systems exists primarily at the molecular level and variability in supramolecular architectures (although shared amongst some amyloid fibrils from different precursor proteins) distinguishes one amyloid protein from another. More recently cryo-EM has contributed significant structural descriptions of the protofilamentous structure of amyloid fibrils. Amyloid fibrils of the SH3 domain, *ex vivo* lysozyme, Insulin, PrP and $ß_2$-m have been the subject of cryo-EM studies (Jimenez et al., 1999, Jimenez et al., 2002, Jimenez et al., 2001, Tattum et al., 2006, White et al., 2009). From paired up to arrangements of six, a variety of protofilamentous arrangements have been observed. Helical repeats are measured in the order of hundreds of Ångstroms whilst axial repeats arising from sub-units along protofilaments are in the order of tens of Ångstroms. Reconstructions of

fibrils combined with inclusion of molecular models of the respective protein typically revealed a cross-ß core where ß-sheets may possess a marginal twist resulting in a long-range fibril twist.

Initially, the direct observation of continuous ß-sheet secondary structure in a core region of Aß demonstrated the resolving power of the cryo-EM technique to probe the structure of amyloid assemblies (Serpell and Smith, 2000). Combined with computational methods of single particle helical reconstruction Sachse *et al.* showed that Aß(1-40) paired protofilaments are laterally displaced with respect to one another forming the helical twist commonly observed in negative stain TEM. Most importantly, inspection of the cross-sections of the three-dimensional density reconstructions of the fibres and comparison to theoretical projections of cross-sections of the cross-ß architecture show a good fit (Sachse et al., 2006) as seen in the concurrently published ssNMR models of the quaternary structure of Aß(1-40) (Petkova et al., 2006). As shown in Figure 1.10a, further analysis unequivocally reveals the protofilaments contain U-shaped structures in a head-to-head helical arrangement, reminiscent of ssNMR models produced of Aß (Luhrs et al., 2005, Petkova et al., 2002). However, there remains some ambiguity related to the exact nature of the protofilament structure possibly due to polymorphic differences in samples studied by different techniques (Sachse et al., 2008).

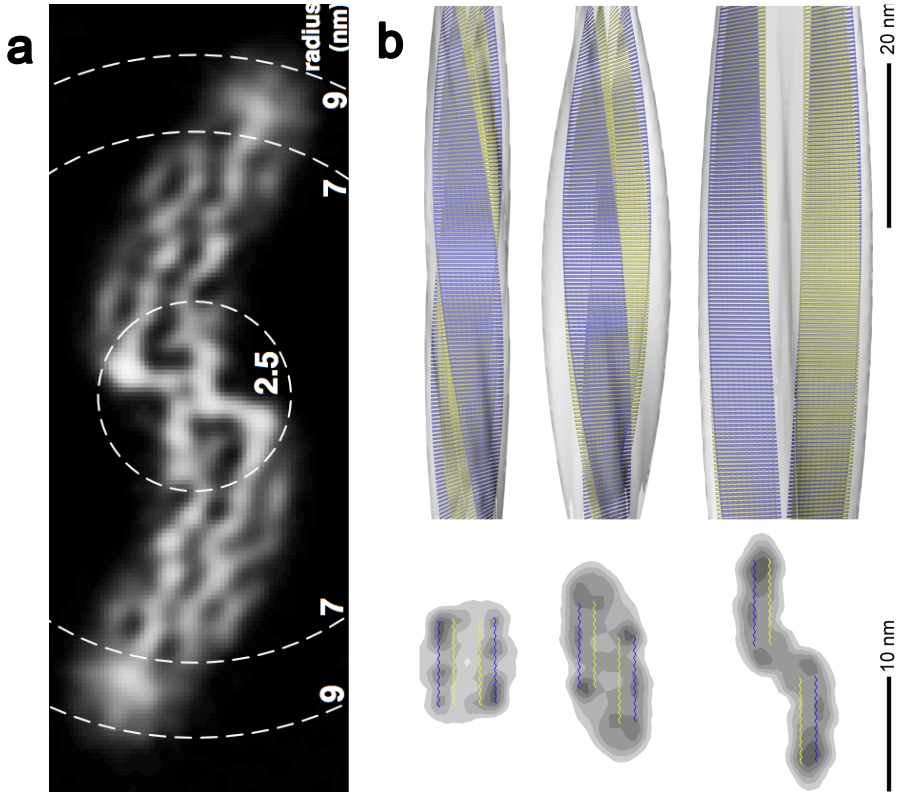


Figure 1.10: A cross section of the reconstructed cryo-EM density of Aß(1-40) at 8 Å resolution, the thin bands of density are assumed to represent polypeptide backbone (a) (image reproduced with permission from (Sachse et al., 2008) Copyright 2008 National Academy of Sciences, U.S.A.). Cryo-EM three-dimensional reconstructions of Aß(1-40) fibril polymorphs, cross sectional slices show the molecular basis for fibril polymorphism of this kind (b) (image reproduced with permission from (Meinhardt et al., 2009)).

Fibril polymorphism is a well-known phenomenon common to self-assembling amyloidogenic peptides and proteins including short fragments of Aß (Colletier et al., 2011), fragments of $ß_2$-microglobulin (Yamaguchi et al., 2005), insulin (Jimenez et al., 2002), IAPP (Goldsbury et al., 2000, Goldsbury et al., 1997), α-synuclein (Heise et al., 2005) and full-length Aß (Paravastu et al., 2006, Petkova et al., 2005). A comprehensive cryo-EM study of fibril polymorphism in Aß(1-40) found a positive relationship between maximum fibril width and helical cross-over distance that can be rationalised to the same model of displaced protofilaments observed before (Meinhardt et al., 2009), also consistent with reports of steric zippers being able to adopt variable modes of packing (Nelson et al., 2005, Sawaya et al., 2007, van der Wel et al., 2007). Figure 1.10b shows the remarkable reconstruction of these fibril polymorphisms and the molecular protofilament structure associated with them (Meinhardt et al., 2009). However, the

most recent cryo-EM work has identified new morphologies yet to be entirely rationalised with our current molecular structural understanding of amyloid from XRFD, SC-XRD and ssNMR (Tattum et al., 2006, White et al., 2009). In these cases the subunits of individual protofilaments are bi-lobed suggesting the presence of multiple protein monomers per axial fibril repeat. This supramolecular arrangement creates a discontinuous ß-sheet structure unlike the majority of models proposed which are continuous.

Although the emphasis is now placed on the role of oligomeric species in toxicity in amyloid related diseases (Irvine et al., 2008) the toxic mechanisms associated with fibrils should not be ignored. A relationship between cytotoxicity and varying fibril morphology has been demonstrated (Petkova et al., 2005, Seilheimer et al., 1997); the stabilisation of the fibrillar state has been suggested as a therapeutic target (Xue et al., 2009) but also, the question has been posed as to what the molecular basis for the prion strains is and indeed fibril morphology may be responsible (Aguzzi and Haass, 2003, Meinhardt et al., 2009, Petkova et al., 2005). Clearly, it is still imperative to understand the structure of the fibrillar state. Combining XRFD and indeed SC-XRD or NMR with cryo-EM can produce models that cover a larger resolution range than each technique individually. It would be of great benefit to obtain more models of amyloid in the fibrillar state to establish whether these features are apparent in all amyloid fibrils. As well as the implications for disease, the technological and methodological developments associated with these investigations represent striking advances in structural biology broadly applicable to the field.

1.8.2 Nuclear magnetic resonance spectroscopy

Solution state nuclear magnetic resonance (NMR) spectroscopy is not able to resolve the structure of molecules above a certain molecular weight and so its application to the study of amyloid fibrils is limited (McDowell and Schaefer, 1996). Conversely, solid state NMR (ssNMR) has proven invaluable where it is not subject to this limitation. Being able to measure internuclear distances up to 6 Å with ^{13}C labelling, relative torsion angles within the polypeptide and the absolute orientation of bonds relative to the magnetic field direction a number of models of amyloids have been reported (Tycko, 2000). This method, unlike the all atom models of crystallography, records a set of distances that when taken together can be combined to computationally find a model that satisfies all of the constraints recorded.

A number of amyloid models in the solid state have been reported in the literature. The structure solved for ccß-met (see Figure 1.8d) was done so with XRFD in conjunction with ssNMR (Steinmetz et al., 2008). The structural determination of a fragment of the amyloidogenic protein transthyretin was also enabled by ssNMR. The fibrillar peptides positioned in amyloid fibrils were found to possess such great regularity, long-range order and limited variation in dihedral angles that they are comparable to a degree of order only found in crystals (Jaroniec et al., 2004). This affords the efficient formation of inter backbone amide hydrogen bonding parallel to the fibre axis, and thus strength.

Currently our most comprehensive models of full-length Aß come from ssNMR data, where models have been produced for Aß(1-40) (Bertini et al., 2011, Paravastu et al., 2008, Petkova et al., 2002) and Aß(1-42) (Luhrs et al., 2005) by several groups. Figure 1.11 shows one such model with ß-strands arranged in parallel, with in register ß-sheets arranged in anti-parallel that perpetuate along the fibre axis (Luhrs et al., 2005).

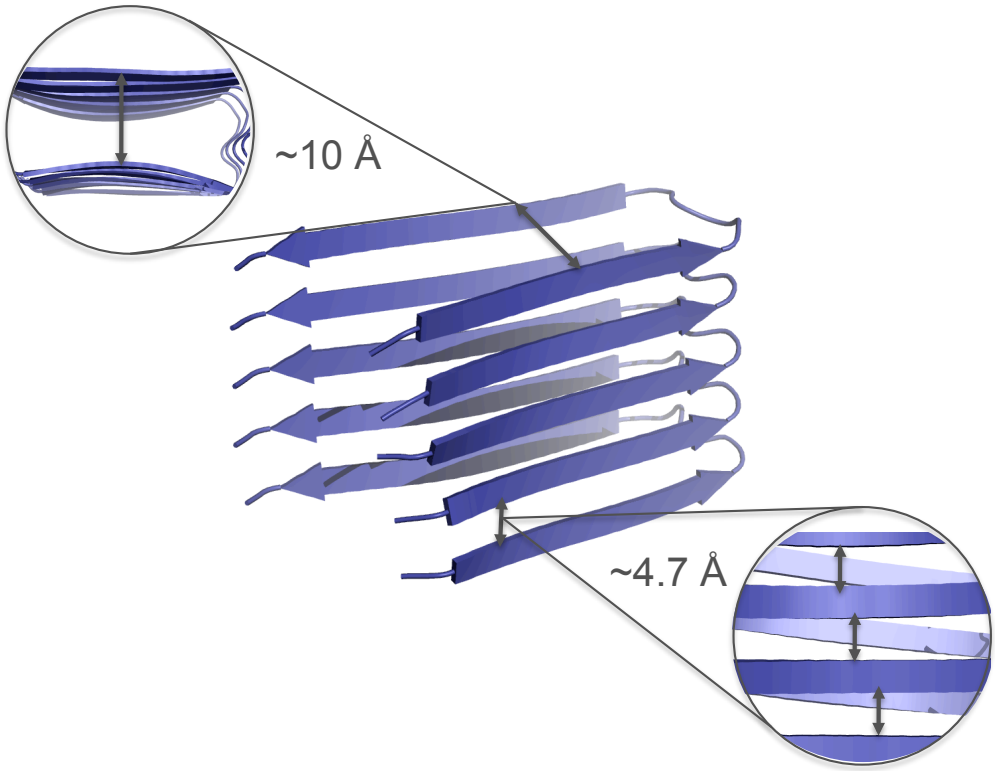


Figure 1.11: A solid state NMR model showing the structure of the Aß(1-42) protofilament exhibiting the characteristic 10 Å sheet and 4.7 Å ß-strand spacing (Luhrs et al., 2005). Graphics generated in PyMol (DeLano, 2002).

1.8.3 X-ray crystallography

Contrary to the widely accepted notion of an inherent inability to crystallise fibrillar assemblies and produce samples suitable for SC-XRD, a set of short peptides have been used in crystallisation studies and used to derive atomic resolution models for amyloid assemblies. Initially the growth of microcrystals produced samples only suitable for XRFD and powder X-ray diffraction (P-XRD) (Balbirnie et al., 2001) and subsequently high-resolution electron diffraction was able to produce unrefined models for a short peptide (Diaz-Avalos et al., 2003). Crystals large enough for SC-XRD have since revealed precise atomic resolution structures describing the fibrillar structure of short peptide sequence (GNNQQNY) derived from the yeast protein Sup35 (Nelson et al., 2005). This model possesses a cross-ß spine structure very similar to the structure proposed for Aß fibrils. These short peptides formed fibrils with ß-sheets within

a single protofilament running in parallel and in register with one another, as is proposed for Aß fibrils. The authors comment that the ß-sheet fibrils are stabilised by hydrogen bonding and opposing ß-sheet protofilaments are able to interact with each other by way of interdigitated side chains, a phenomena named the steric zipper.

The steric zipper, shown in Figure 1.12a, describes the side chains of individual ß-strands nestling between those of ß-strands of adjacent ß-sheets. This close interaction of interdigitating side chains creates a water free environment (Nelson et al., 2005) on one side of the peptides that suggests how fibrils of amyloid interact to create higher order structures. A second crystal packing for GNNQQNY is shown in Figure 1.12b. It is reasonable to postulate that a crystal structure represents an energetic minima for a system, and as such the discovery of a second isoform illustrates the adaptability of the cross-ß architecture. Sawaya *et al.* later characterised other short segments from various amyloid-forming proteins, depositing eleven crystal structures representing these, as summarised in Table 1.2 (Sawaya et al., 2007).

These structures can be classified based on their overall ß-sheet architecture and the relative orientations of these. This classification led to the model shown in Figure 1.13 describing eight classes of steric zippers, three of which remain to be observed experimentally. However, it should be remembered that these structures are of crystals and so represent a very different energetic space to fibrils, which are possibly kinetically trapped. Indeed it has been highlighted that some differences do exist between crystals and fibrils of GNNQQNY (Lewandowski et al., 2011, Marshall et al., 2010). With the example of the short peptide GGVVIA derived from Aß, the authors demonstrate that *in vitro* GGVVIA (Aß(37-42)) interacts with native Aß (Sawaya et al., 2007) and so conclude the peptide might represent a native amyloid structure. Interestingly, when an overlapping sequence from Aß, MVGGVV (Aß(35-40)), was subjected to the same experimental method it formed a different structural class with ß-strands arranged in anti-parallel. This observation led the authors to hypothesise that fully matured Aß fibrils may adopt more than one conformation. Either in the sense that entire fibrils may conform to just one of

the many conformations available in a heterogeneous population of fibrils or somehow single fibrils adopt many conformations along their length (Sawaya et al., 2007). Certainly the packing of Aß fibrils in many different ways to form polymorphic fibrils does in part explain the difficulty encountered in the attempts to structurally classify full-length Aß fibrils.

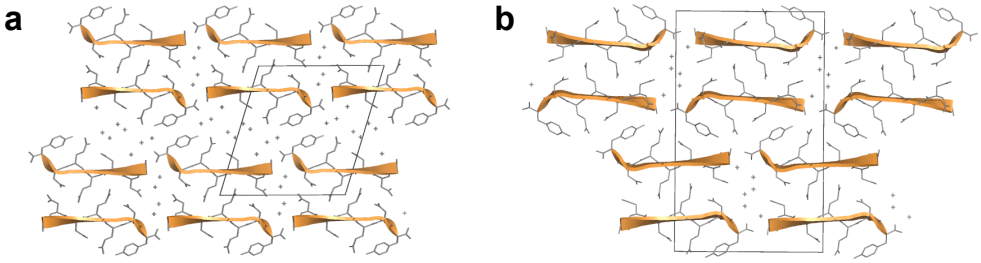


Figure 1.12: The structure of GNNQQNY in its two crystal forms packed into a monoclinic cell deposited as 1YJP (a) and orthorhombic cell as deposited in 2OMM (b). Graphics generated in PyMol (DeLano, 2002).

Protein (organism)	Sequence	Length	a	b	c	α	β	γ	Space Group	PDB ID
Aß (human)	GGVVIA	6	16.372	40.094	4.746	90.00	90.00	90.00	P 21 21 2	2ONV
	MVGGVV (1)	6	9.699	15.851	25.862	74.69	86.93	77.18	P 1	2ONA
	MVGGVV (2)	6	15.148	9.576	23.732	90.00	96.90	90.00	P 1 21 1	2OKZ
Insulin (human)	LYQLEN	6	9.666	28.003	17.346	90.00	96.24	90.00	P 1 21 1	2OMP
	VEALYL	6	18.425	9.613	21.975	90.88	96.12	100.93	P 1	2OMQ
Rnase (bovine)	SSTSAA	6	42.008	4.830	12.888	90.00	90.00	90.00	P 21 21 21	2ONW
Sup35 (yeast)	GNNQQNY (1)	7	21.937	4.866	23.477	90.00	107.08	90.00	P 1 21 1	1YJP
	GNNQQNY (2)	7	4.934	23.324	37.548	90.00	90.00	90.00	P 21 21 21	2OMM
	NNQQNY	6	21.153	4.870	23.130	90.00	102.93	90.00	P 1 21 1	1YJO
	NNQQ (1)	8	4.854	16.014	15.546	90.00	96.91	90.00	P 1 21 1	2ONX
	NNQQ (2)	8	15.479	4.915	30.552	90.00	90.00	90.00	P 21 21 21	2OLX
Tau (human)	VQIVYK	6	4.863	61.926	15.413	90.00	98.11	90.00	P 1 21 1	2ON9
Prion (human)	SNQNNF	6	14.002	4.879	15.100	75.23	75.88	78.89	P1	2OL9

Table 1.2: Crystal structures of short segments taken from amyloid forming proteins.

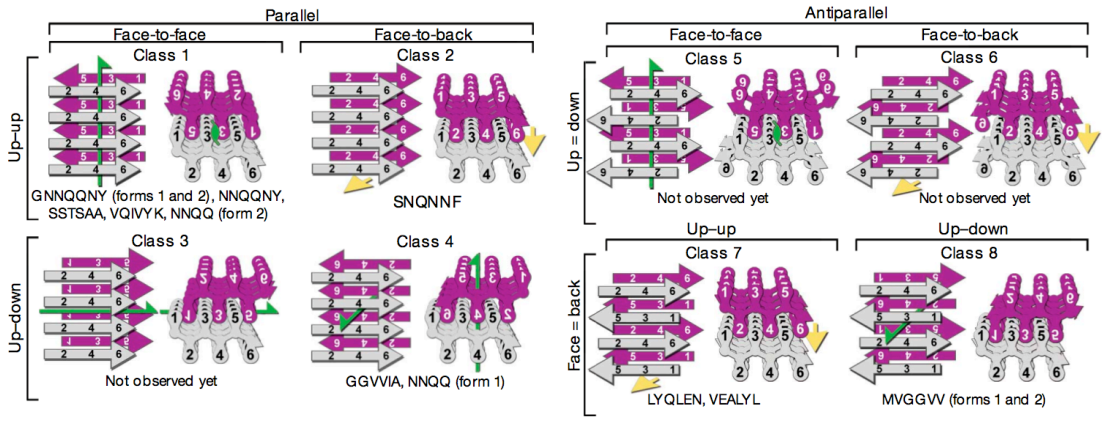


Figure 1.13: A schematic representation of the architectures formed by synthetic short peptide. Reproduced with permission from (Sawaya et al., 2007).

1.9 Self-assembly – From natural to designer

In addition to the uncontrolled, undesirable self-assembly described, it is incredible that in a controlled environment the self-assembly process can be functional and useful. Macromolecular structures on the order of nanometres are frequently required in biological and now man-made systems. Synthesis of such structures as single units or by irreversible polymerisation is impractical and so a route to the production of such structures is available through self-assembly. The self-association and assembly of proteinaceous material is a recurring theme in biological systems. Non-covalently linked assemblies can be formed from small molecules with the molecular weights of large polypeptides to as low as 450 Da for low-molecular weight gelators.

A number of examples can be described to illustrate the breadth of native uses of such a fibrillar system and the structural basis for their formation. Figure 1.14 shows examples of self-assembled fibrillar systems utilised by nature and inspired by nature but designed by man. Such systems all exploit the anisotropic nature of these assemblies for their respective functions, which can be broadly categorised as providing support for larger structures or making up the effective parts of protein motor systems. A self-assembled fibril is an ideal structure for a cellular support. Its mechanical strength is derived mainly from a high density of lateral and longitudinal non-covalent bonds in the body of the fibril and has the ability to de/repolymerise permitting the reorganisation of the structures that it supports. At the same time structural support can be provided on much shorter length scales by coiled-coil fibrils for instance to create dimerisation interfaces for transcription factor heterodimerisation (Glover and Harrison, 1995) or even catalytic protein domain dimerization (Joshuator et al., 1995). The motif responsible for coiled-coil formation has been rationalised to a set of design rules that have allowed this conformation to be used to design useful materials but systems that are *de novo* designed are also now reported in the literature with a range of possible uses (Boyle and Woolfson, 2011).

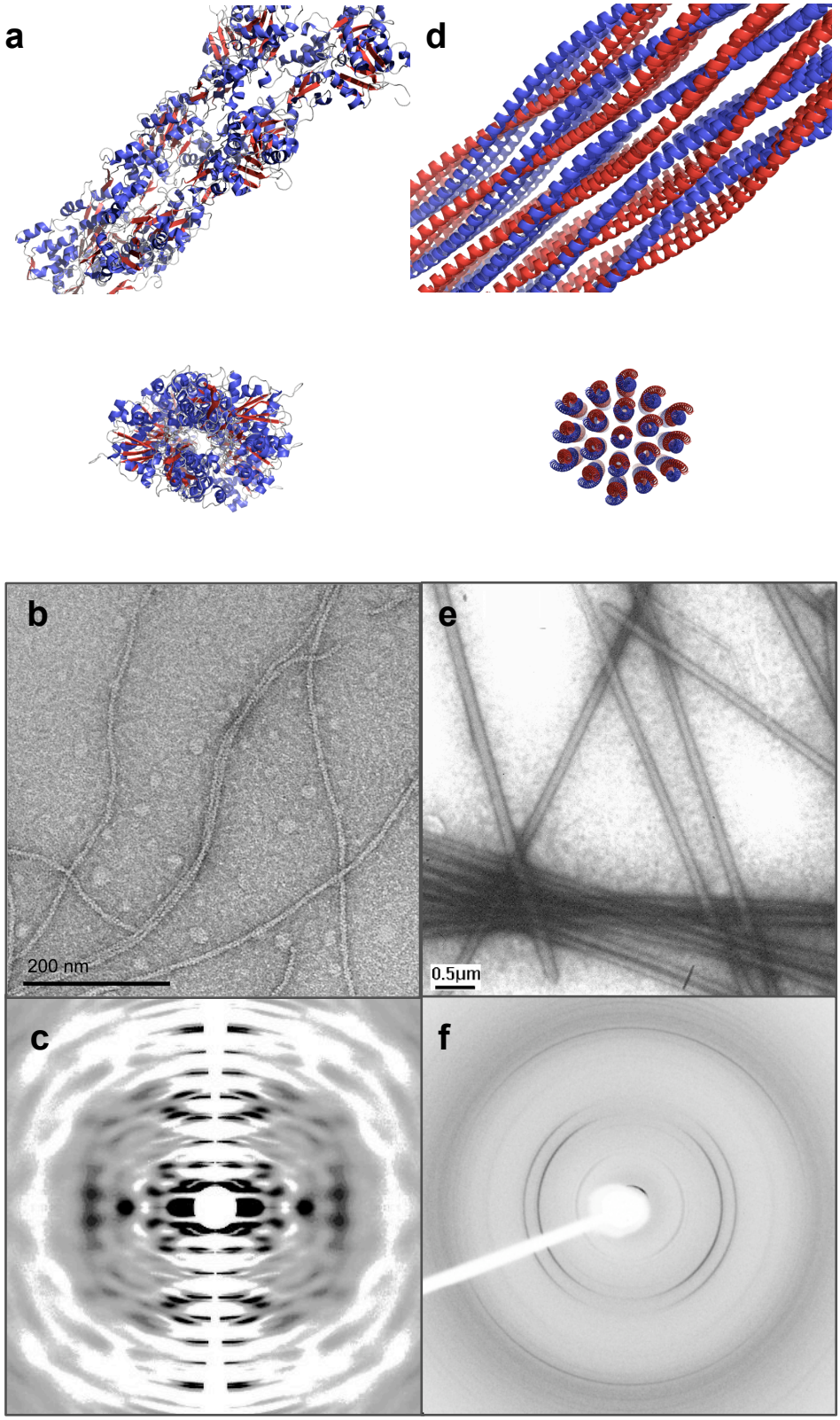


Figure 1.14: From natural to designer: A range of self-assembled systems. Fibrillar assemblies may also be formed by mixed secondary structural elements as in actin (a-c) (Oda et al., 2009) or α-helical coiled-coils as in self-assembled fibres (SAFs) (d-e) (Papapostolou et al., 2007) though still X-ray fibre diffraction has been imperative in investigations of these systems. All fibre axes are vertical. Figure adapted from (Morris and Serpell, 2010).

The current understanding of these systems is in part gained from X-ray fibre diffraction studies. The following sections expand on these examples, illustrating the use of fibrils in nature, to being inspired by nature and designed from the bottom up.

1.9.1 Actin

Elements of the cytoskeleton that provide structural support but also motile forces to eukaryotic flagella and cilia include filamentous microtubules. Actin filaments, as shown in Figure 1.14a, play a vital role in the formation of filopodia during neuronal growth. Microtubules and actin filaments consist of self-associated monomers of tubulin and actin respectively. Tubulin and actin have guanidine and adenosine triphosphate hydrolytic activity that gives rise to conformational states favouring self-assembly. Actin exists in two forms: G-actin, the globular form and F-actin, a polymerised fibrillar form (Straub, 1942). An increase in salt concentration drives a transition from G-actin to F-actin (Straub, 1942) and in this manner assembly can be controlled and reversed. The crystal structure of G-actin in the adenosine triphosphate (ATP) and adenosine diphosphate (ADP) bounds forms were determined in 1990 by Kabsch *et al.* (Kabsch et al., 1990). Based on the structural determination of F-actin by TEM a model was proposed for F-actin by fitting the crystallographic atomic model of Actin-ADP into the helix of the low-resolution constraints (Holmes et al., 1990). In a reiterative computational approach the F-actin model was refined further starting from G-actin in an attempt to better match the experimental fibre diffraction and represent the that changes that might occur in the fibrillar transition of G to F-actin (Lorenz et al., 1993). This represents a multi-technique approach that reflects the need for corroborative data in XRFD analysis. This model has subsequently been accepted and used successfully to describe additional molecular features of actin in muscle fibres (Lorenz et al., 1995, Vibert et al., 1997). Using the same method further refinements have been made (Holmes et al., 2003) but most significantly in the advent of technical advances in fibre diffraction and sample preparation new diffraction data at higher resolution has recently been reported. Combined with computational refinement the new data has been used to generate

a new model of F-actin from the originally G-actin crystallographic model (Oda et al., 2009). The authors note differences in the new F-actin model that result in greater stabilising contacts between monomers but an inherent instability associated with the monomer conformations themselves, perhaps giving structural insights into actin dynamics.

1.9.2 Coiled-coils

Perhaps in stark contrast to the majority of self-assembling systems designed and studied to date are systems based on α-helices. Coiled-coils were first described by Pauling and Corey (Pauling and Corey, 1953a) as α-helical structures that can laterally associate via non-covalent interactions to form oligomers from dimers to hexamers, as shown in Figure 1.14d. The structural motif found in many biological systems such as α-keratin and myosin and the rules that govern the final assembly are relatively well understood. Thus helical dimers were designed to assemble in a staggered manner, the assembly mechanism promotes the formation of repeating a heterodimeric unit that produces fibres with lengths in the order of several hundred micrometres (Pandya et al., 2000). These assemblies are referred to as self-assembling fibres (SAFs) and the principles of their design are to align interacting amino acids in the positions that are responsible for dimer interface. The interface is based on the 1-ABCDEFG-7 heptad consensus and the inclusion of amino acid with interaction potential at positions A, D, E and G (Pandya et al., 2000). Other reports of 'sticky ended' coil-coils can be found producing fibrils of a diameter of about 50 nm with pronounced striations along their fibre axes corresponding to an α-helical configuration (Papapostolou et al., 2007). In line with the desire to be able to modify the molecular features of a monomer to change the properties of the final self-assembled material variations of the basic SAF design have been engineered to produce hydrogels. By maintaining the residues responsible for dimerisation interface but modifying the remaining residues specific lateral fibre interactions can be produced based on weak hydrophobic and hydrogen bonding interactions. These changes produce a new class of hydrogelating SAFs (Banwell et al., 2009).

1.9.3 *De-novo* designed hydrogelating systems

Self-assembling short peptides not related to any native protein sequence have been engineered based on a set of design principles – side chains attracted via hydrophobic forces, aromatic residues for **π-π** interactions, electrostatic interactions via oppositely charged residues and alternating hydrophilic-hydrophobic residues to make the peptide molecule asymmetrical in its attraction to water (Aggeli et al., 1997). The resulting designer peptides successfully assemble in water, from monomers into self-supporting ß-sheet rich tapes producing thermostable hydrogels. Further work has demonstrated the ability to design gelating self-assembling materials that are tunably and responsive to external stimuli as summarised by Figure 1.15.

Using amino acids that have a high propensity for forming ß-sheets and including a sequence designed to form a type-II ß-hairpin a self-assembling hydrogel is formed (Rughani et al., 2009, Schneider et al., 2002). The use of high ß-sheet propensity amino acids and the positioning of the ß-hairpin signature affords an impressive amount of control and potential for modifying material properties but the choice of the amino acid sequence is key to factors stabilising these structures. A repeating $[lys\text{-}val]_n$ motif produces an amphiphilic molecule, resulting in increased lateral association. Derivations of the basic peptide design (Figure 1.15) have rational effects on material properties whereby the assembly is controlled by altering pH and temperature.

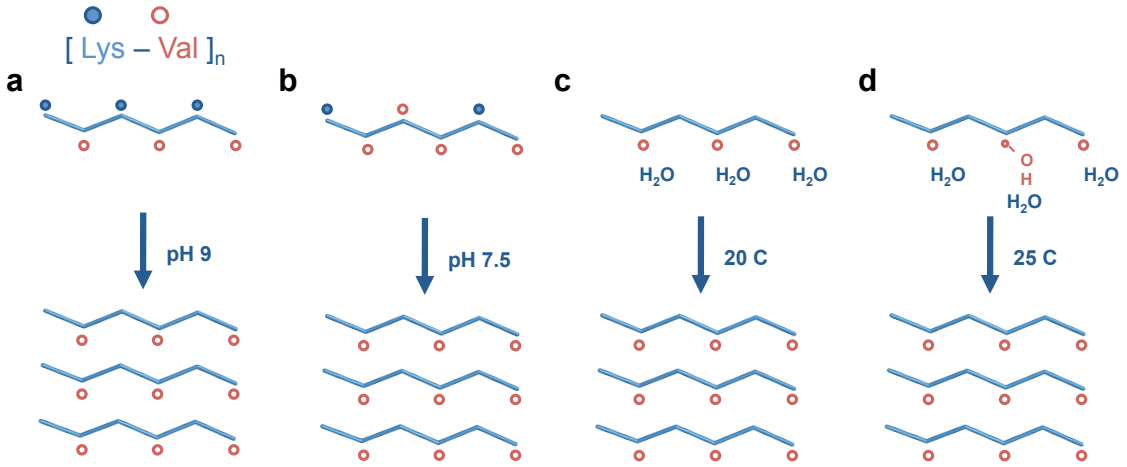


Figure 1.15: Self-assembling peptide design principles (Schneider et al., 2002). The pH of assembly may be controlled by charged residue mutations (a-b) and the temperature of assembly controlled by hydrophobicity changes (c-d).

The pH at which assembly occurs can be modulated by modifications to the hydrophilic side of the molecule (Figure 1.15a-b), mutations to hydrophobic residues result in the removal of charge-charge repulsion and thus self-assembly occurs at lower and importantly more physiological pHs. Conversely the temperature at which assembly occurs can be modulated by modifications to the hydrophobic side of the molecule. Self-assembly normally requires increasing temperature to cause the desolvation of hydrophobic residues. Thus mutations to less hydrophobic residues on the hydrophobic side of the peptide results in a system that self-assembles at higher temperatures (Figure 1.15c-d) due to the less readily desolvating residues (Rughani and Schneider, 2008).

1.10 Self-assembly design and prediction

1.10.1 Design

The work by Schneider *et al.* described previously is an example of the control available in a designed self-assembling system. To the scientist trying to synthesise a large macromolecular system there are a number of strategies available (Whitesides et al., 1991) but a self-assembling system has a number of appealing features. The technical challenge of creating a relatively

small monomer is much less than the final assembled structure and where the assembly is stabilised by non-covalent forces is potentially controllable. Additionally where the self-assembly process must be spontaneous and represent a thermodynamically favourable transition the generated structures occupy a minima making them natively robust (Whitesides et al., 1991). The challenge in making such a system lies in designing the self-assembly mechanism, where currently the principles of non-covalent forces are well understood in isolation but the application of these is an emerging science.

The assembly interfaces that have been designed to date have been simple and effective (Schneider et al., 2002) or, when complex, based on naturally occurring mechanisms (Pandya et al., 2000, Papapostolou et al., 2007). Thus there is a drive to be able to predict self-assembling systems in nature not only to identify new systems with which to characterise this phenomenon but also with the hope of identifying new design rules that can in the future be applied to useful man made systems. For a comprehensive review of current prediction methods see Section 5.1.

1.11 Concluding remarks

The investigations described highlight the value of using short model peptides as a means for elucidating the mechanisms by which amyloid forms and is stabilised. It has proved difficult to obtain high-resolution structures of full-length amyloids and so the structure of an amyloid fibril may be determined by using a combination of structural descriptions of many shorter peptides arrived at by different techniques. Although high-resolution amyloid structures exist for a range of short-peptides (Sawaya et al., 2007), the concern over whether these fully represent the structure in the fibrillar phase remains an open question (Marshall et al., 2010). Certainly though they are one of the valid viewpoints from which amyloid self-assembly is characterised. Other techniques have successfully described this architecture at the molecular length scale including ssNMR (Petkova et al., 2002, Petkova et al., 2006) and XRFD (Makin et al., 2005, Sikorski and Atkins, 2005, Sikorski et al., 2003). These structures collectively represent

common amyloid motifs and allude to the driving forces behind amyloid fibril formation. By taking different sequences from full-length amyloids, core sequences possibly important in a polypeptides propensity to be amyloidogenic have been identified. As well as enabling a greater description of this structural conformation, investigations of this sort reveal design principles of self-assembly and may be used in the design of useful self-assembling systems (Schneider et al., 2002).

2 Introduction II:

Low-molecular weight gelators

2.1 Introduction

The self-assembly phenomena of peptidic molecules has been described in detail with a view to understand disease but also the rational design of fibrillar self-assembled materials. These self-assemblies may in some cases also undergo another phenomena, known as hydrogelation (Banwell et al., 2009, Rughani et al., 2009, Schneider et al., 2002). Due to the potential uses of these in material science, the understanding of self-assembly also lends itself to dissecting the phenomena of hydrogelation as described in the following sections.

2.2 Hydrogels and their applications

Hydrogels consist of a matrix that immobilises water to form a self-supporting gel. These gels have a high water content and so, if formed under physiological conditions, are thought to be biocompatible. Hydrogels can be formed in a number of ways; all based on the principle of creating scaffolds that trap solvent, with a wide variety of precursors available to form the hydrogel scaffold. The molecules that can be used may be based on standard assembling polymers, macromers (preformed supramolecular structures) or self-assembling systems. Consequently, a great number of hydrogel materials exist, that can be broadly described as being synthetic or non-synthetic, depending on the nature of their precursor. Further classification may describe hydrogels as either polymeric or molecular. In the former, hydrogelators are preformed and subsequently crosslinked, whereas in the latter case hydrogelators are monomeric molecules with self-assembly potential (Xu et al., 2007) such that the gel matrix may be formed spontaneously as shown in Figure 2.1.

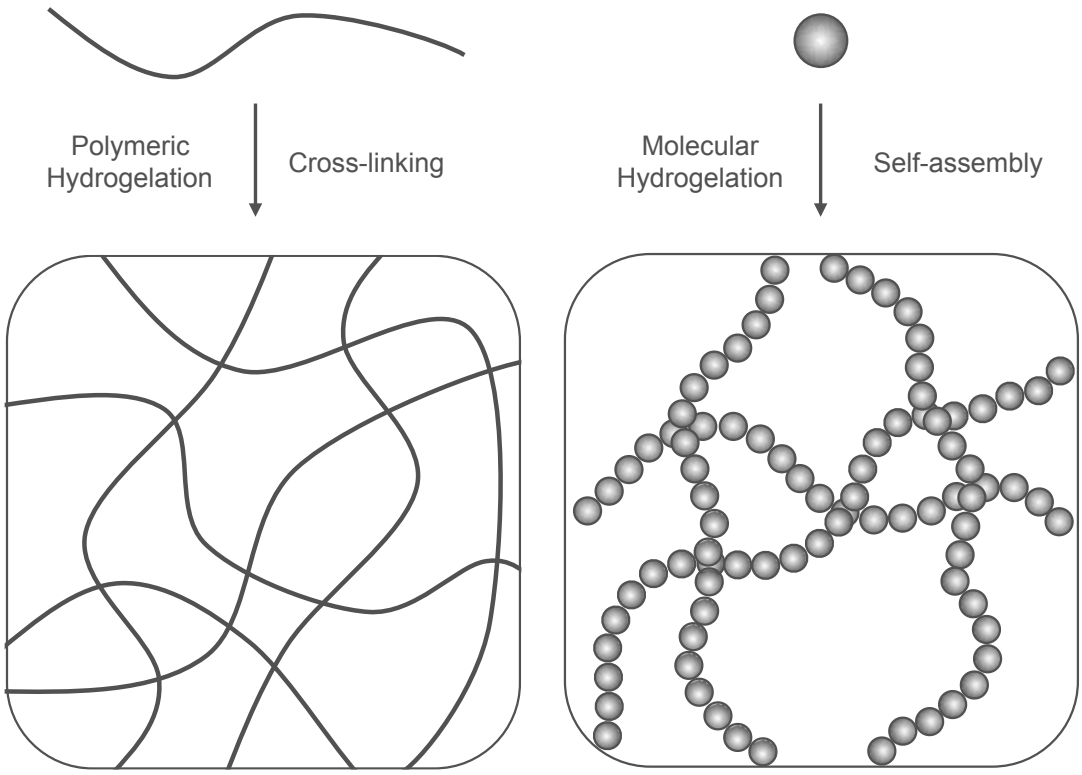


Figure 2.1: Polymeric vs molecular hydrogelation. Adapted from (Xu et al., 2007).

The macromers used for polymeric hydrogels may be polysaccharides or synthetic polymers, but their detailed discussion is beyond the scope of this work (for review see (Peppas et al., 2009)). In brief, the biological macromers reported in the literature have included agarose, alginate, chitosan, hyaluronan, fibrin and collagen. Synthetic macromers, used in tissue engineering, have included polyethylene glycol, polyvinyl alcohol, and polyacrylates including polyhydroxyethylmethacrylate. In these cases, the hydrogelation of the solvent the macromers are dissolved in is induced by activating cross-linking reactions between the fibrillating material. These well-established polymer/macromer based hydrogelating systems are mechanistically and mechanically dependent on the cross linkages between the fibrils of the gel matrix. Indeed, where cross-linking can be specifically controlled in hydrogelating systems, this appears to be the key factor to controlling material properties. These cross-linkages may be based on chemical bonding (covalent or ionic bonds), weak interactions (hydrogen bonding) or permanent and temporary physical entanglements (Peppas et al., 2009).

There have been many classes of gelators reported in the literature. For a comprehensive review, a large proportion of these systems are discussed in detail elsewhere (Adams, 2011, Peppas et al., 2009). The polymeric hydrogels that have been produced are typically very stable, an attribute somewhat desirable in certain applications; however for a material to act as a cellular scaffold it needs to be degradable to allow the cells to eventually produce their own extracellular matrix (ECM) to substitute the hydrogel matrix. Cells have indeed been successfully cultured in polymeric hydrogels but due to the low degradability of these materials, cell viability decreases over time (Elisseeff et al., 2000). In addition, the chemistry involved in polymerising synthetic monomers (often free radical based) is toxic to cells (Bryant et al., 2000). Moreover, the agents required for chemical cross-linking raise the same concerns regarding safety, whereby the control of these reactions typically requires the use of toxic transition metals (Peppas et al., 2009).

In addition to circumventing these problems, it is recognised that there is a need to be able to understand the physiochemical and structural determinants that lead to hydrogelation such that reproducible and reversible hydrogels can be produced with prescribed structures and properties (Kopecek, 2009). This has been achieved, to an extent, with pre-existing gelating systems but an alternative route to the production of molecular hydrogels is through the use of a low-molecular weight gelator (LMWG), in particular discussed here peptidic LMWG. These remove the necessity for harmful chemical agents as they spontaneously self-assemble and laterally associate in a stimuli responsive manner and, due to the non-covalent interactions that stabilise these systems, may presumably be dynamically controlled and even reversed. A large body of research has emerged describing the development of self-assembling systems that are synthetic molecular hydrogelators (Adams, 2011, Ulijn and Smith, 2008). These often consist of a short peptide sequence conjugated to an aromatic moiety such as fluorenylmethyloxycarbonyl chloride (Fmoc) or naphthalene. The sections herein describe these systems.

Molecular hydrogels are already in use and established as biomaterials; some examples of research in this field have been translated into a clinical setting with applications ranging from the manufacture of contact lenses to implantations used in cosmetic surgery (Kopecek, 2009). Hydrogels also have a great deal of prospect for applications such as tissue engineering and molecular scaffolds, afforded by their similarity to the ECM. They could be used to culture a particular tissue type and then facilitate delivery to the required patient's transplantation area. To this effect, hydrogels have been used to culture chondrocytes (Jayawarna et al., 2006) and hydrogels with varied mechanical properties have been shown to exhibit preferential support of specific cell types (Jayawarna et al., 2009). It is well known that cells have mechotransduction systems and will respond appropriately to the different matrices they are proliferating in (Ingber, 2006). To this effect it has been demonstrated that mesenchymal stem cells chosen cell lineage may be temporarily determined by chemical induction factors but is ultimately dependent on matrix elasticity (Engler et al., 2006). Further increasing the compatibility of hydrogels with tissue engineering, cell adhesion motifs may be engineered into gelating molecules. Sites for cell adhesion within a gel matrix should change the spatial environment that culturing cells can inhabit and would be vitally important where cellular responses to spatial tension cues are required for effective growth. Examples include laminin, fibronectin, collagen and Arg-Gly-Asp (RGD) cell adhesion motifs (Peppas et al., 2009). The RGD cell adhesion motif, in particular, has been reported as having been used in hydrogel biomineralisation applications (Hartgerink et al., 2001); moreover, hydrogels containing this motif were able to encourage cell adhesion of dermal fibroblasts and promote their proliferation (Zhou et al., 2009).

Recent work has promised to expand this repertoire with innovative and novel applications. Numerous examples exist: hydrogels have been used in the treatment of simulated uranium wounds (Yang et al., 2005) and as new kinds of anti-inflammatories (Yang et al., 2004). Hydrogels may be used for the immobilisation of enzymes affording hyperactivity and increased stability (Wang et al., 2008, Wang et al., 2007) but also for use in enzyme activity assays and thus additionally reporting on inhibitor performance (Yang and Xu, 2004). Low but

appreciable inherent conductivity has been demonstrated in hydrogels formed from Fmoc-Leu-Leu-Leu (Xu et al., 2010) and the potential for light harvesting materials has been demonstrated by the incorporation of fluorophores into naphthalene conjugated dipeptides showing assembly dependent fluorescent resonance energy transfer (FRET) (Chen et al., 2010b). The photonic-electronic properties of hydrogels is becoming an increasing area of interest where hydrogels of Fmoc-Phe-Phe have been shown to have properties similar to quantum dots (Amdursky et al., 2010). Finally, recent work has now proven the catalytic benefits of Pt coating high surface area / low volume ratio templates such as tobacco mosaic virus, hydrogels could act as similar templates thus having applications as transition metal high surface area catalysts (Gorzny et al., 2010).

These applications benefit from the presence of the gel matrix and its structural constituents but another set of applications in drug delivery are conceptually based on another property of hydrogels. Diffusion is considered the principle mode of solute transport through gel materials (Amende et al., 1995) and the dependency of dye propagation on molecular weight through hydrogels has been explored (Sutton et al., 2009). A hydrogel could be considered to have size exclusion chromatographic properties but has the unusual property of retaining high molecular weight species for greater lengths of time where these molecules become trapped in the interfibre water reservoirs. The nature of these reservoirs presumably defines diffusion properties. The radii of the reservoirs in one hydrogel were estimated by determination of proton transfer throughout the gel and the reservoirs size shown to be dependent on the concentration of the gelator (Amdursky et al., 2009). With an apparent ability to control the factors thought to affect small molecule diffusion rates within hydrogels, these materials have naturally been considered as candidates for time-dependent release of small molecules. Indeed, a tetrapeptide has recently been shown to immobilise the anti-cancer drug Doxorubicin, subsequently facilitating slow release of the compound into a physiological environment, indicating a potential use in future drug delivery systems (Naskar et al., 2009). It is recognised that the release of any small molecule, including a drug, could be tuned by modification of the

gel network density and mesh size or rapid release via disassembly of the hydrogel network (Adams, 2011).

Clearly hydrogels have huge potential in bio and nanotechnological applications. The fundamental insights required to generate and develop these applications is ever pressing. Yet the ability to characterise and understand these self-assembling systems has broader implications for understanding any self-assembled system, may it be found in a materials science application or in understanding diseases caused by the aberrant self-assembly of a protein or peptide.

2.3 Material properties

Hydrogels display a variety of responses when subjected to external mechanical forces and this response defines how close to a solid or fluid the material is. A solid material's response to external mechanical forces may be to deform plastically or elastically under an applied stress. In the former, energy input into the system enables bond breaking and the material does not recover; but in the latter case, an elastic material will regain some of its original form where the energy input into the system is stored. A liquid-like material will exhibit elastic behaviour but also viscosity that is a degree of flow. Given time, a polymeric liquid - such as a hydrogel - will flow under a slow deformation, defined as its viscosity. So too if an external force is applied to a polymeric liquid for a period of time shorter than for it to reach a new structural equilibrium then, on release of the deforming force, the material will regain at least some of its original form, defined as its elastic property. A polymeric liquid possesses both of these characteristics and so is termed to be viscoelastic (Lin, 2003 pg 49). The study and characterisation of these material properties is called rheology. Rheology concerns the application of an external force to a material (stress) and the measurement of the materials resulting displacement as a response to this (strain) (Zbigniew K, 2002). A purely elastic materials strain will be perfectly in phase with the applied stress where one can imagine a perfectly elastic material returning to its original

form instantly. In contrast, an ideal viscous material will exhibit a phase difference of 90^{o} between the applied stress and strain response (Fried, 2003 pg 208-9). A great number of methodologies for rheological measurement exist from capillary and couette to cone-and-plate and parallel-plate rheometry (Fried, 2003 pg 461-8) to sophisticated laser tweezer microrheometry (Veerman et al., 2006). In the same manner, a great number of material parameters may be measured, but with respect to hydrogels it is useful to report the two values G' and G". The definitions of these values are numerous and often interchangeable. Briefly, G' is the storage modulus and represents the elastic properties of a viscoelastic material and G" is the loss modulus, representing the viscous properties of a material where this value reflects the energy lost from the system as heat (Lin, 2003 pg 63). Both G' and G" values are measured in Pascals. A liquid would be expected to have very similar G' and G" values (Chen et al., 2010a) but as it becomes a gel the G' value will increase above the G" value. Hence G' indicates gel strength by how many orders of magnitude it is above the G" value.

It is important to be able to understand the viscoelastic properties of a hydrogel and indeed control them, as previously described, as in tissue engineering applications these properties have a direct effect on the cells growth. The physical and structural properties of gels that determine matrix elasticity: cross-linking, mesh size and gel strength must all be inherently linked. The degree of cross-linking may affect the gel strength but concurrently increased cross-linking per unit length of matrix fibrils should decrease pore size thus also influencing diffusion properties. Greater control over any one of these interdependent material properties is beneficial, where mesh/pore size has a direct bearing on solute transport through materials when doing so by diffusion (Pappenheimer, 1953) but has also been implicated selectivity by certain cell types (Zeltinger et al., 2001).

2.4 Peptidic based low-molecular weight hydrogelators

As described in Chapter 1, a number of studies have aimed to identify the minimal peptidic sequences that are able to self-assemble. In many cases, this type of work has identified short hexapeptide sequences from larger polypeptides (Maurer-Stroh et al., 2010, Thompson et al., 2006). Remarkably though, this process has identified even more minimalist sequences, the shortest of which to date reported in the literature is the Phe-Phe motif found in the 42-residue polypeptide Aß (Gazit, 2002). Systems designed to model amyloid self-assembly have since been designed from this short motif and characterised revealing insight into self-assembly structure (Makin et al., 2005). The aromaticity of these sequences is presumably responsible for a hydrophobic collapse resulting in amorphous aggregation but the ability of the aromatic moieties to π-π stack and the backbone to hydrogen bond introduces directionality and anisotropy into the self-assembly process. The importance of aromatic moieties in amyloidogenic proteins has been emphasised in a number of independent reports in the literature (Gazit, 2002, Makin et al., 2005), in particular, where under the correct conditions diphenylalanine alone will self-assemble into vesicles and peptide nanotubes (Reches and Gazit, 2006). Coincidentally, a number of LMWG gelators have emerged with this same assembly motif, notably Fmoc-Phe-Phe (Smith et al., 2008). However, the addition of aromatic conjugations such as Fmoc removes the necessity to have an assembly competent peptide sequence as evident by early reports of gelation from Fmoc-Ala-Ala (Zhang et al., 2003) and Nap-Gly-Gly (Xu et al., 2007). The use of conjugate groups like Fmoc and Nap also introduces potentially useful properties such as well characterised fluorescent excitation/emission profiles and the opportunity to design FRET mechanisms (Chen et al., 2010b). Further, where the assembly is no longer driven solely by the sequence motif, assembly can be controlled by exploiting the physiochemical characteristic of the peptidic sequence.

A number of self-assembling LMWG with high aromatic content are currently the subject of intense investigation because of their potential applications in nanotechnology, synthetic

biology and tissue engineering. These molecules are peptidic, typically with two amino acid residues N-terminally conjugated to aromatic moieties. On a nano to mesoscopic level their self-assembly creates a three-dimensional network or scaffold that on the macroscopic level forms a self-supporting hydrogel.

Figure 2.2 shows four classes of LMWG based on the conjugation of a dipeptide to an aromatic moiety, in these cases either Fmoc, naphthalene, pyrene or spiropyran. The potential variation of the aromatic moiety and conjugated dipeptide sequence produces a rich diversity in potential LMWGs. With such a large number of potential gelators, the literature reports a huge range of gels having been produced for a variety of applications as previously described. Their discovery is often serendipitous and so questions emerge that if answered could potentially expand the repertoire of these molecules to more designed and informed applications. In short these are, how does the self-assembly mechanism of these molecules work and how is this related to hydrogelation? Further, how might the structure of these systems relate to the matrices formed and thus the material properties governed by these matrices?

As previously eluded to, given the large range of physiochemical properties possessed by amino acids and conjugated aromatic groups there is an opportunity to engineer the self-assembly mechanism of LMWGs to be stimuli responsive through control of non-covalently linked polymerisation. Designing supramolecular assemblies that are based on these interactions, to be able to control one interaction is to be able to control the assembly and disassembly of the entire system. The use of these interactions in supramolecular assembly certainly affords control over this process, however a distinction should be made between controlling self-assembly and hydrogelation because some systems have been observed to be pre-assembled prior to gelation (Chen et al., 2011) and in other systems self-assembly is a necessary prerequisite and codevelopment to gelation (Chen et al., 2010a) as discussed in Chapters 7 – 8.

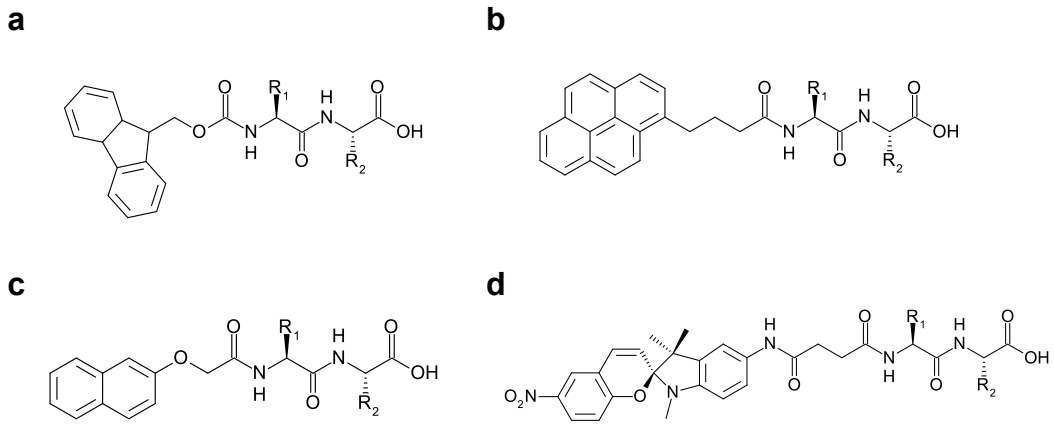


Figure 2.2: Four classes of peptidic low-molecular weight gelators classified by their aromatic conjugation to FMOC (Zhang et al., 2003) (a), Pyrene (Zhang et al., 2004) (b), Naphthalene (c) (Xu et al., 2007) or Spiropyran (d) (Qiu et al., 2009). Figure adapted from (Adams, 2011).

From the summary in Figure 2.3, it is clear that the individual strength of each interaction is relatively weak (a covalent bond is in the order of several hundred kJ/mol/bond in strength) but with the sum of each interaction very stable assemblies can be produced. Hydrogen bonding is extremely directional and typical bond distances are expected to be 2.8 Å (Corey and Pauling, 1953) whilst π-π stacking interactions are reported to be variable and dependent on geometry (Meyer et al., 2003); a survey of 30, 444 pairs of aromatic interactions in proteins however finds the centroid to centroid distance of aromatic interactions to be ca. 5 – 6 Å (McGaughey et al., 1998). Electrostatic interactions are even less directional, weak and additionally their existence is highly dependent on their environment. Hydrophobic interactions are the weakest and least geometrically constrained of all the non-covalent interactions (Mart et al., 2006). These non-covalent forces are central to the control of LMWG self-assembly, and in isolation are relatively well understood but the fully informed and predictable application of these to supramolecular self-assembly is arguably an emerging science (Whitesides et al., 1991). The balance between hydrophobicity, π-π stacking interactions and hydrogen-bonding is still not fully understood and a matter of some debate (Johnson et al., 2011) but the following sections describe a variety of methods used to control supramolecular assembly, many of which are based on manipulation of the weak chemical interactions described above.

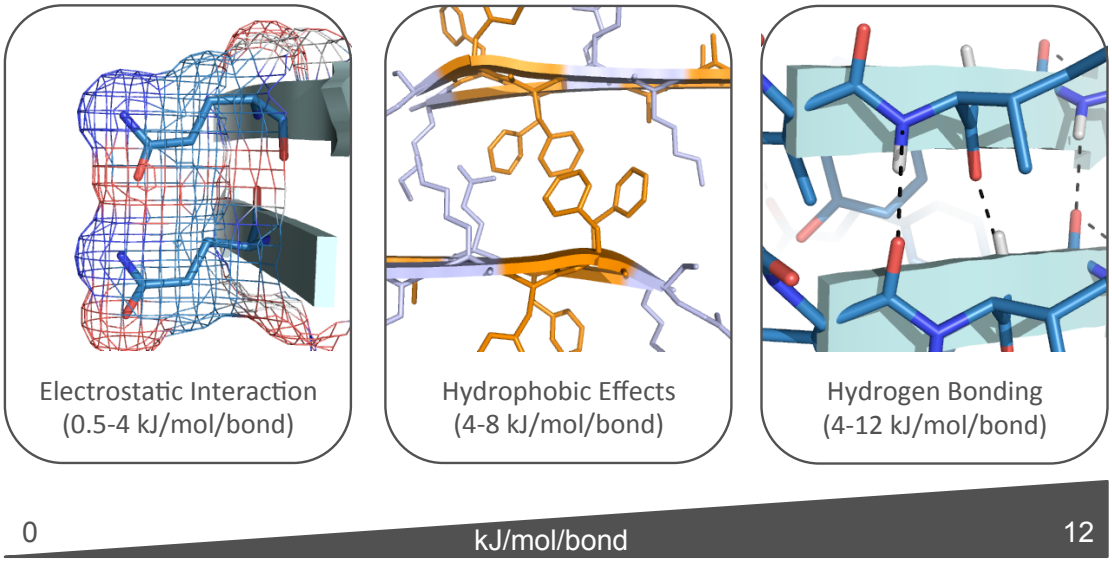


Figure 2.3: The order of strength of typical non-covalent stabilising interactions, from electrostatic interactions (Sikorski and Atkins, 2005), through to hydrophobic effects (Makin et al., 2005) to hydrogen bonding. Bond strengths taken from (Mart et al., 2006).

2.4.1 Enzymatic control

One route to hydrogelation is to take an assembly competent LMWG and prevent self-association by steric hindrance and charge repulsion. As such, hydrogelation has been enzymatically induced by hydrolysis of C-terminal esters on Fmoc dipeptides. This is achieved by cleavage of a phosphate group from tyrosine by alkaline phosphatase (Yang et al., 2005) and shown in Figure 2.4. Furthermore, reversible hydrogelation has been achieved with this system based on amide bond formation (reverse hydrolysis) using thermolysin leading to self-assembly and disassembly controlled by the esterase subtilisin (Das et al., 2008). Further examples of enzyme controlled self-assembly where a dipeptide sequence is enzymatically linked to an Fmoc group have been reported (Toledano et al., 2006).

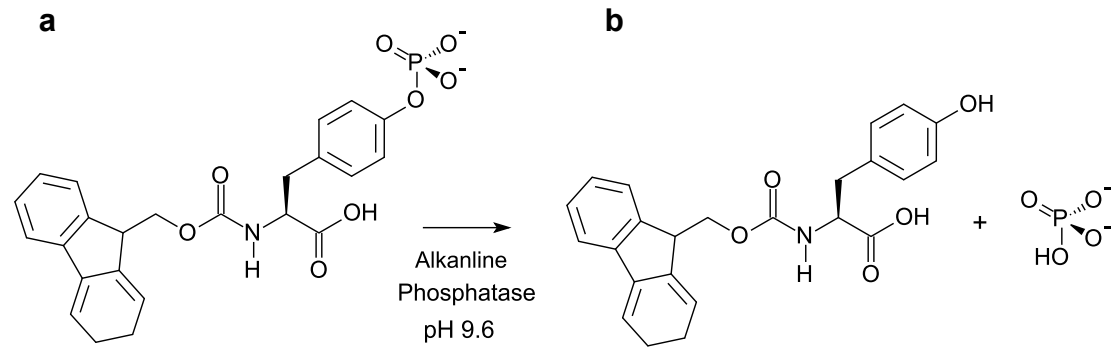


Figure 2.4: The dephosphorylation of Fmoc-Tyr. The phosphate group is found at the γ position of the Tyr residue (a) and is removed by alkaline phosphatase results in the less bulky and neutral molecule that is self-assembly competent (b). Figure adapted from (Yang et al., 2005).

2.4.2 Ligand binding control

A ligand may be used to control gelation and has been reported for an Fmoc system where gelation only proceeds after vancomycin binding (Zhang et al., 2003). Demonstrating the versatility of this control mechanism in another dipeptide system a similar ligand-binding site for vancomycin disassembles the supramolecular structures of the gel matrix and disrupts the hydrogel (Qiu et al., 2009). By choosing this control system, not only is assembly controllable but gelation may be used as a detection system or to enable a response that is only observed when the correct stimuli is present or absent.

2.4.3 UV/Vis Light control

A unique example of hydrogelation control based on the photo-sensitive group Spiropyran (SP) has been reported in the literature. The advantageous feature of this molecule is that it can be isomerised between a non-planar SP form and the planar merocyanine (MC) form by exposure to visible and UV light respectively creating a bulky group that would sterically hinder supramolecular assembly. The SP conjugate reported by Zhang *et al.* and the conversion between the non-planar and planar isomers is shown in Figure 2.5 (Qiu et al., 2009).

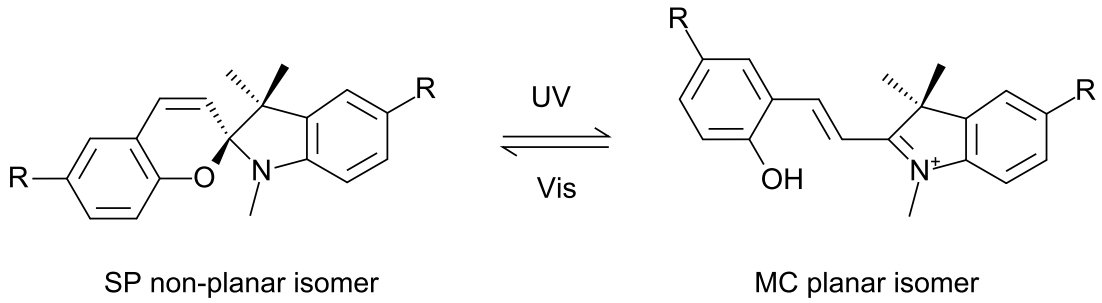


Figure 2.5: The isomerisation of Spriropyran (SP) in its non-planar form to the planar Merocyanine (MC) isomer by exposure to light. Adapted from (Qiu et al., 2009).

In the MC planar isomer, the system is able to make π-π interactions whilst the non-planar isomer does not exhibit this behaviour, in this manner the authors ascribe self-assembly to be controlled via aromatic interactions (Qiu et al., 2009). In addition to these effects, steric hindrance presumably also plays a role in assembly control where the isomers are structurally quite different it can be envisaged that only one isomer is compatible with the packing arrangement necessary for self-assembly.

2.4.4 pH control

Figure 2.6 shows that aromatic conjugated dipeptides exist in equilibrium between the carboxyl (COOH) and carboxylate (COO^-) forms. As such, a pH drop can be used to induce self-assembly and gelation where carboxylate peptide termini may prevent self-assembly via charge-charge repulsion. Typically, to induce gelation, acids such as HCl have been used to lower the pH of a dipeptide solution but this creates localised heterogeneously distributed areas of low pH resulting in hydrogels with irreproducible rheological properties. It has been shown that through the use of Glucono-δ-lactone (GdL) a homogenous pH change can be induced leading to hydrogels of greater strength and reproducibility (Adams et al., 2009). GdL spontaneously hydrolyses in water to *D*-Gluconate and hydrogen ions. The rate of dissolution of the GdL is much faster than that of the rate of hydrolysis and hence the pH decreases homogeneously, not only does this result in higher quality gels and but also has the benefit of permitting the monitoring of the self-assembly and gelation process.

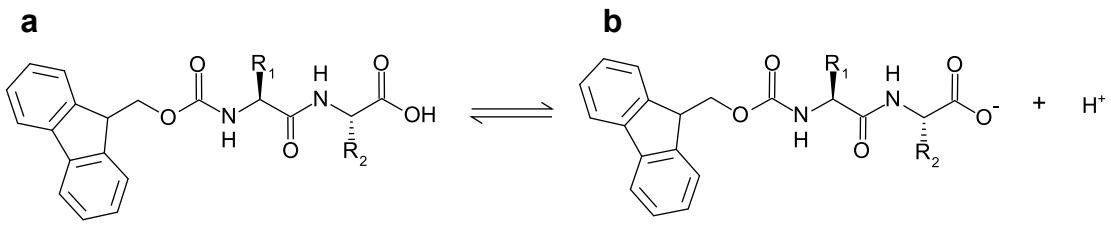


Figure 2.6: N-terminally aromatic conjugated dipeptides exist in equilibrium between the neutral protonated form and the ionised form. At pHs lower than the carboxyl groups pKa the equilibrium shifts left and at high pHs right. Adapted from (Tang et al., 2009).

The reaction products of the hydrolysis of GdL are in fact quite complex, resulting in other products in addition to *D*-Gluconate and hydrogen ions including gluconic acid and glucono-γ-lactone all in equilibrium (Pocker and Green, 1973); the kinetics of the pH drop and final pH reached are however, directly proportional to the concentration of GdL, and highly reproducible (Adams et al., 2009).

2.4.5 Electrical protonation

Using the same mechanism of self-assembly control via the lowering of pH, the C-termini of conjugated dipeptides may be protonated electrically. Using a gold electrode grid suspended in a solution of hydroquinone and the LMWG, Fmoc-Leu-Gly, it has been shown to be possible to induce thin film gelation across the gaps of the electrode grid by applying a current across the system such that the hydroquinone is oxidised to 1,4-benzoquinone releasing hydrogen ions. This creates a local pH drop below the pKa of the dipeptide inducing self-assembly and gelation (Johnson et al., 2010).

2.4.6 Temperature control

It is well known that the solvation of hydrophobic residues is temperature dependent; a hydrophobic residue under the right conditions can be solvated. If two compounds are considered, **1** highly and **2** weakly hydrophobic, **1** will be desolvated at a lower temperature

than **2,** where greater energies/temperatures are required to desolvate **2** due to affinity for water (see Chapter 1 and Figure 1.15). If desolvation is a factor that can be utilised to control self-assembly and hydrogelation then these phenomena can be incorporated into the LMWG design, as has been shown for longer designer self-assembling peptides (Rughani et al., 2009). Indeed such an assembly control mechanism has been elegantly demonstrated with the use of Fmoc derived LMWGs with a variety of hydropathy indices as shown in Figure 2.7. The three systems were noted to gel as they were cooled (Vegners et al., 1995). However the critical concentration above which gelation can occur was much greater for the most hydrophilic system, Fmoc-Ala-Asp, demonstrating that at the same temperature, to overcome the solvation shell a degree of molecular crowding is required.

2.4.7 Mechanical control

A unique example of mechanical control over gelation has been reported in the literature where cyclic LMWG have been shown to exist as a sol in solution when at quiescent conditions but respond to mechanical shear forces by gelating. The systems exhibit thixotropic behaviour, that is to be viscous under normal condition but exhibit lower viscosity upon agitation, which could simply be accounted for by gel instability but it is interesting that the initial hydrogelation of these systems requires mechanical agitation (Xie et al., 2009).

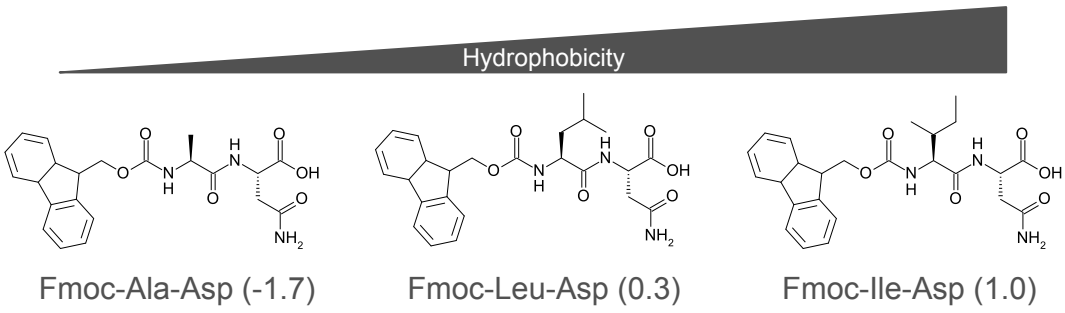


Figure 2.7: Three Fmoc based LMWG with increasing hydropathic indices. Hydropathy index values calculated by summation from (Kyte and Doolittle, 1982) based on the dipeptide sequence alone.

2.5 The protonation assembly mechanism and pKa

The assembly of aromatic conjugated dipeptides via protonation through the lowering of pH has been extensively investigated by Adams *et al.* and Ulijn *et al.* and is discussed in more detail in the following section. The assembly mechanism through this change in environment has been shown to be inherently linked to the pKa of the C-terminus of the dipeptide of the LMWG. The self-assembly of the system Br-Nap-Ala-Val has been monitored using thioflavin-T fluorescence and it was found that assembly inferred by a fluorescence intensity increase only occurs when the pH reaches 5.4, correlating well with the apparent pKa of the carboxy-terminus (Chen et al., 2010a). The pKa of the carboxy-terminus in this case is surprisingly high where for a traditional protein or peptide it might be expected to be ~3.7 (Ellenbogen, 1952) but this phenomenon can be explained for many peptidic LMWGs. The system Fmoc-Phe-Phe is found to self-assemble at physiological pH but this has been rationalised to be due to perturbation of the pKa of the carboxy-terminus to ~6.4 allowing assembly at similar pHs (Tang et al., 2009) although it should be noted that the Phe-Phe motif probably also makes significant contributions to the assembly potential of this system. Clearly when assembly and gelation occurs through this mechanism, the pKa plays an important role and where it has been noted that the pKa of C-termini can shift dramatically in peptidic LMWG, investigations have been conducted to measure these shifts across a range of systems. It has been found that there is a direct correlation between predicted partition coefficient (clogP: the logarithm of the predicted ratio of the concentrations of solute in octanol and water) of the whole molecule, a measure of hydrophobicity, and the experimentally determined apparent pKa. In a library of naphthalene conjugated peptides, pKas were measured at values ranging between 4.0 – 6.8 (Chen et al., 2010c).

It is perhaps no surprise that in a LMWG the hydrophobicity of such a small molecule and thus the hydration would have dramatic effects on the pKa of nearby ionisable groups. It is useful to consider this phenomenon as a result of a changing environment around the ionisable group in

question. It is well established that the pKa of ionisable groups are found to be dependent on the nature of the environment that the group is present in. When a polypentapeptide sequence ([Ile-Pro-Gly-Val-Gly]$_n$) is engineered to contain a single Asp residue it is possible to measure the pKa of that group depending on the neighbouring amino acid residues. The theoretical pKa of this group of 3.71 (Lide, 2004) but is observed to shift to 10.0 when in close proximity to multiple Phe residues whilst when Phe residues are engineered as far away from the Asp group as possible whilst retaining the peptide structure the pKa value is 6.7. Urry *et al.* attribute this phenomena to competition of hydrophilic and hydrophobic residues for hydration resulting in dramatic changes to the pKa of the ionisable groups competing for the hydration shell (Urry et al., 1993). Further, in a survey of empirically determined pKa values for ionisable groups within folded proteins the relationship between environment and pKa is clearly complex but three main factors can be attributed to influence ionisable group pKa: charge-charge interactions, charge-dipole interactions (hydrogen bonding) and the Born effect (dehydration) (Grimsley et al., 2009). The pKa's of carboxyl groups increase in environments of low dielectric constant, such as that experienced when shielded from water by molecular crowding, structural burial or proximity to a hydrophobic residue.

Thus as shown in Figure 2.8, pKa, hydrophobicity and concentration of LMWG are inextricably linked. Increasing the concentration of the gelator will increase molecular crowding and no doubt place the ionisable groups of a LMWG into more hydrophobic environments having dramatic effects on their apparent pKa values and in turn affecting the pH at which self-assembly occurs. Although the environment created by the supramolecular assembly of these molecules may have large effects on the assembly mechanism. With this in mind, it is possible to envisage the ability to select or design systems tailored for assembly in particular conditions.

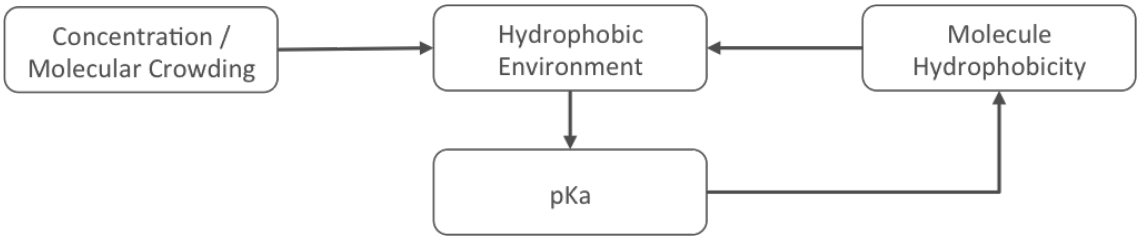


Figure 2.8: The pKa of ionisable groups, in this case carboxyl termini, are inextricably linked to concentration and the hydrophobic affect.

2.6 Current structural insights

As with the study of amyloid proteins the structural determination of the fibrillar material that makes up a hydrogel matrix is not a straightforward process and an additional challenge then remains in how to establish a link between these architectures and macroscopic measured rheological properties of the hydrogel. A number of investigations have used spectral and structural techniques to propose low-resolution descriptions of the architectures adopted by a range of LMWG (Hughes et al., 2011, Smith et al., 2008, Xu et al., 2010). Only a few molecular models for dipeptide LMWG have been proposed and those that exist describe general molecular features as opposed to being comprehensive structural descriptions of the systems they represent. It is clear and generally well accepted what intermolecular forces stabilise these structures; although the relative significance each contributes is a matter of some debate and with a view to control these systems understanding this is key. To date, the models proposed are reasonably based on experimental observation but are not validated against three-dimensional structural data by simulation. The following sections describe current structural understanding of these supramolecular assemblies. Further, the models described draw conclusions about interatomic distances without directional information i.e. data is collected from unaligned samples of gel in wide-angle X-ray scattering (WAXS) measurements. Hence, some caution should be taken in the inspection.

2.6.1 Fmoc based systems

The groups of Xu and Ulijn have conducted a great deal of valuable work on Fmoc conjugated dipeptide systems. Molecular dynamics & WAXS studies of Fmoc-Tyr gels produced by enzymatic-induced gelation (see Section 2.4.1) emphasise the importance of aromatic π-π stacking in the structures from this LMWG. Smith *et al.* proposed a model for Fmoc-Phe-Phe assembled in water with anti-parallel ß-sheets and chirally ordered fluorenyl groups in an anti-parallel arrangement indicated by induced CD signals and a red shifted Fmoc fluorescence spectrum (Smith et al., 2008). Similar biophysical data for Fmoc-Leu-Gly has been collected, showing red-shifted fluorenyl fluorescence data and induced chirality of these groups indicated by CD (Adams et al., 2009). To satisfy these structural observations, a cylindrical fibril model for Fmoc-Phe-Phe was proposed, as shown in Figure 2.9 (Smith et al., 2008). The aromatic stacking of the fluorenyl groups is said to be in a J-aggregate geometry (see Section 3.2.5) due to the observation of a red-shifted fluorescence peak (Smith et al., 2008). Distances are assigned to the ß-strand separation at 4.3 Å, although these distances are made by non-directional WAXS measurements.

The systems Fmoc-K(Boc)LV and Fmoc-VLK(Boc) containing the KLV motif reported to be taken from the Aß polypeptide have been shown to have very different nano-scale and macroscopic material properties. The former assembles to create gels with greater strength, which is ascribed to structural differences in the assemblies whereby the stronger gel former (Fmoc-K(Boc)LV) consists of unorientated multiple stacks of ß-sheets (Cheng et al., 2010) and it is inferred the weaker gel former (FmocVLK(Boc)) does not and instead consists of more ordered stacks of ß-sheets. Though interesting, no further suggestion of the connection between the structure and material properties is offered.

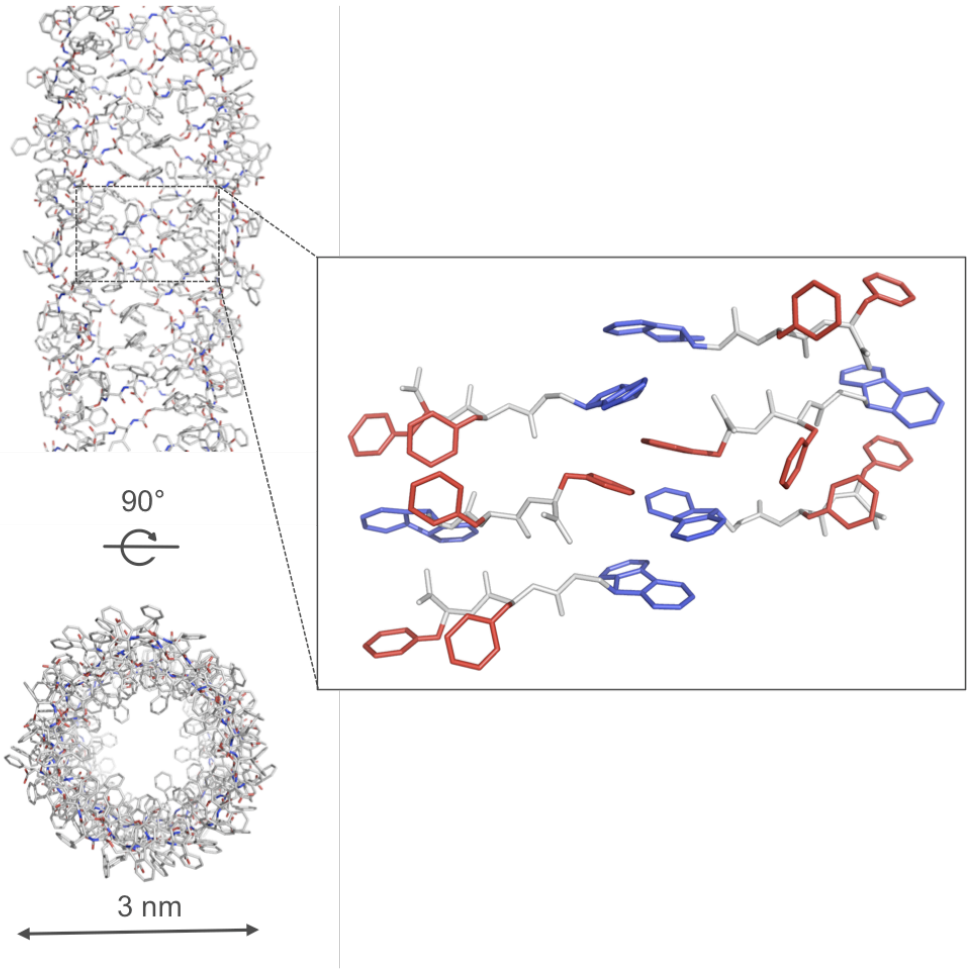


Figure 2.9: A model of Fmoc-Phe-Phe in the fibrillar gel form, as reported in (Smith et al., 2008). In the expanded view Fmoc groups are coloured blue and Phenylalanine groups red.

Other investigations have reported nanotubular structures for Fmoc based systems in the gel phase. Xu *et al.* observed structures of fibrillar Fmoc-Leu-Leu-Leu with an inner core diameter of 7 nm and outer diameter of 16 nm. They attributed these structures to an underlying nanotubular architecture consisting of antiparallel ß-sheets with a hydrogen bonding distance of 4.6 Å, an intersheet distance of 11.5 Å and π-π stacking distances of 3.6 – 3.8 Å (Xu et al., 2010). Fluorescence spectra indicated electron delocalisation, which may be attributable to π-π stacking interactions. The proposed structure is shown in Figure 2.10, where it can be seen that the fluorenyl groups appear to be stacked in a direction approximately parallel to the peptide backbone hydrogen bonding which is spaced at 4.6 Å (Xu et al., 2010). It is not clear how the reported distance between fluorenyl rings is shorter than the ß-strand spacing and yet both appear to be aligned to the same axis. Evidently, some care should be taken with these structural

models where they may require further refinement, however, they do show unique structural features that may prove invaluable for identifying structure function relationships in these systems in the future.

Figure 2.11 shows another model proposed for Fmoc-SF-OMe. Reported to be stable by molecular dynamics, the dipeptide length-repeat is on average 16.2 Å, with ß-strands spaced at 4.9 Å and a fluorenyl separation of 3.7 Å. Hughes *et al.* ascribe the 10.1 Å in the WAXS data to the repeat length of Fmoc dipeptide pairs but lack directional diffraction data to fully support this conclusion. The fluorenyl-fluorenyl separation is assigned as 3.7 Å extending approximately parallel to the fibre axis when the ß-strands are separated by 4.9 Å (Hughes et al., 2011). As well as being interpreted from unaligned WAXS data, in a similar manner to other reports, it is difficult to rationalise how the two unique separations can be aligned to the same axis if the molecules are also approximately aligned to the same fibre axis.

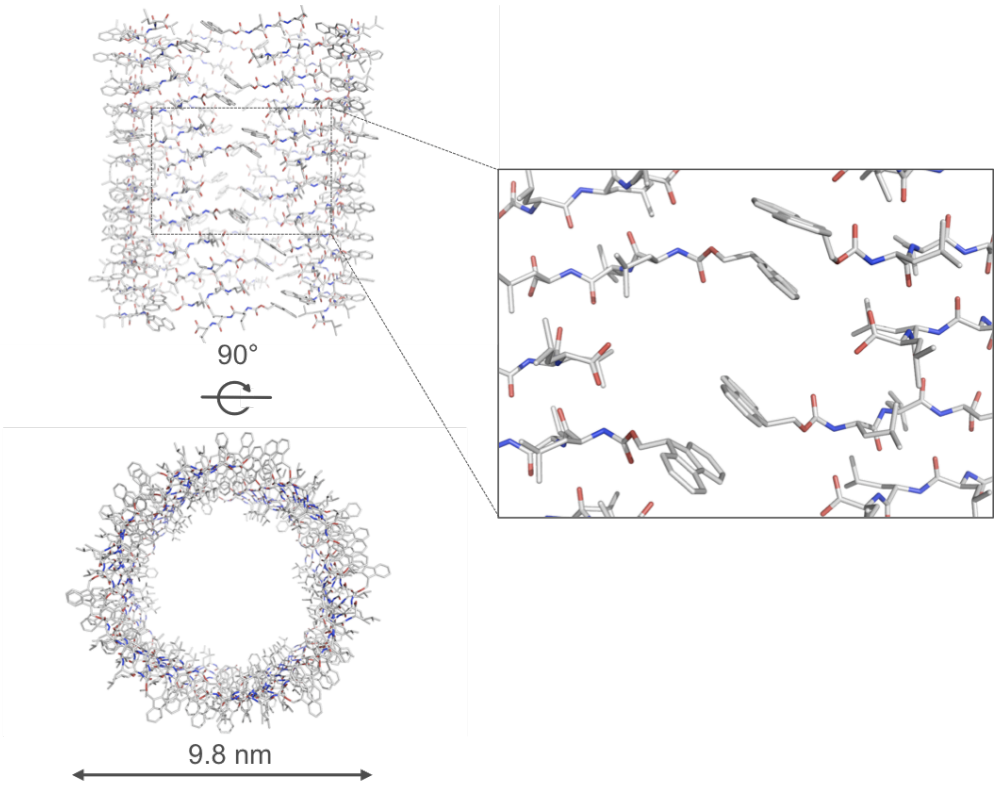


Figure 2.10: The proposed fibrillar model of Fmoc-Leu-Leu-Leu in the gel phase, as reported in (Xu et al., 2010). PDB co-ordinates kindly provided by Dr. Andrew Smith (University of Bayreuth).

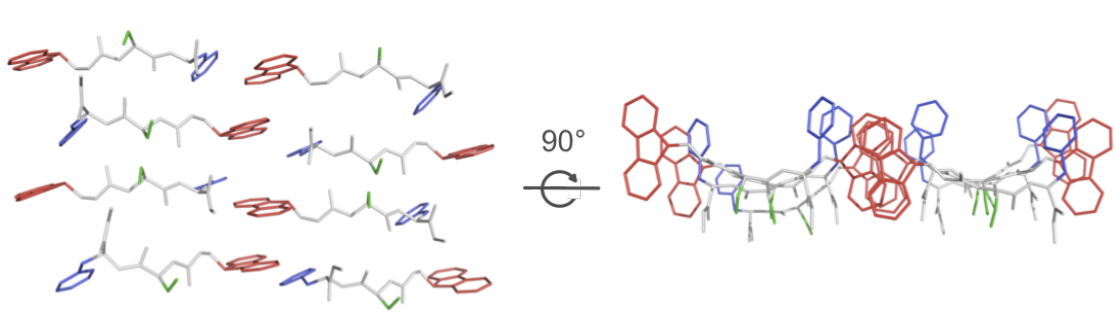


Figure 2.11: Fmoc-SF-OMe structure, as shown reported from (Hughes et al., 2011). Fmoc groups are coloured red, serine green and phenylalanine blue. Atomic co-ordinates kindly provided by Pim Frederix and Prof. Rein Ulijn (University of Strathclyde).

These studies place great emphasis on the adjacent stacking of fluorenyl groups but less emphasis is placed on the importance of hydrogen bonding in the reports of the proposed structures. In truth, it is probably the case that each of the stabilising interactions in these structures are in careful balance and so when one is disrupted, where self-assembly still occurs, the others are compensating. Investigations into other different but related LMWG have in part addressed these concepts.

2.6.2 Naphthalene based systems

Two systems recently reported present a unique opportunity to investigate hydrogen bonding in these systems (see Chapter 8). 1-Nap-Gly-Ala & 1-Nap-Ala-Gly are both found self-assemble under acidic conditions, however the former system is a gel former with eventual crystallisation over time but the latter system crystallises *in situ* enabling structure determination by XRCD. It was found that it was possible to predict the structure of the crystal forming dipeptide computationally, as shown in Figure 2.12, but this method does not effectively predict the structure of the gel phase indicating a possible structural discrepancy between the crystal and gel phases (Adams et al., 2010a). Despite this, the structures predicted for 1-Nap-Gly-Ala displayed hydrogen bonding patterns subtly distinct from the crystalline phase hinting that this non-covalent interaction could be key to fibre gel phase formation giving some insight as to

how slight modifications to the amino acid sequence of a dipeptide affects hydrogen bonding patterns and thus whether an assembly competent system gels or crystallises.

Further work on naphthalene-conjugated dipeptides has in part addressed the importance of π-π stacking interactions in these structures. By making substitutions at the 6-position of the naphthalene it might be possible to interfere with the electronic structure of naphthalene such that π-π stacking interactions are perturbed. A brominated naphthalene dipeptide has been reported to retain assembly potential despite the theoretical withdrawal of electrons from the π orbital electron system of the naphthalene by the more electronegative bromine atom (Chen et al., 2010a) as discussed in Chapter 7 and 8. This may be a case where when one intermolecular force is disrupted and, where gelation still occurs, other forces are compensating but it may also indicate a greater importance for the role of hydrogen bonding in the stabilisation of these systems.

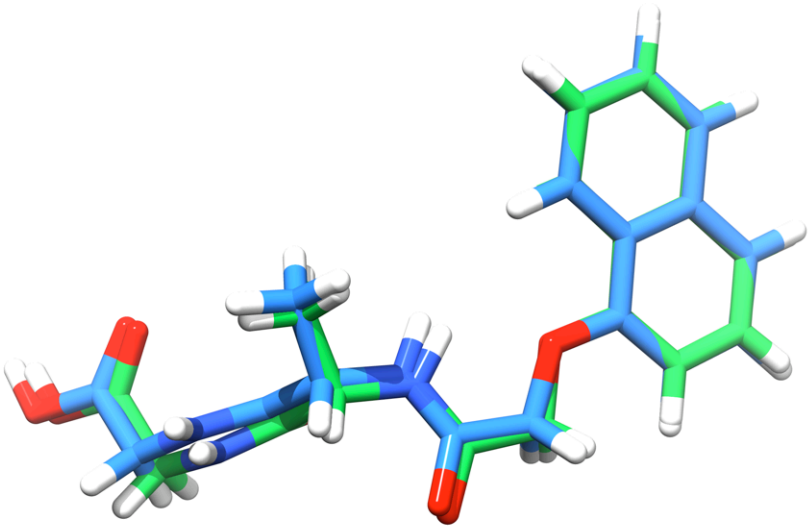


Figure 2.12: An alignment of the crystal structure of the naphthalene conjugated dipeptide 1-GA with a computationally predicted structure for the same molecule (Adams et al., 2010a). Graphics produced in Chimera (Pettersen et al., 2004).

Taken together these studies might indicate an important role for hydrogen bonding in determining the nature of the assembly formed, whether crystalline or gelled. Clearly the hydrogen-bonding pattern will have a very immediate effect on whether a structure formed is one-dimensional and thus fibrillar but from the observations made on Fmoc based systems the importance of π-π stacking interactions are also evident. A true structural understanding of these systems and deconstruction of the balance between stabilising interactions can only come from a comparative analysis of a range of systems all molecularly related but subtly different. No attempts to do this to date have been reported, nor has an attempt to systematically correlate structural features with gel properties been reported. Recently, the effect of molecular structure on molecular physiochemistry and gel material properties has been reported for a library of naphthalene-conjugated dipeptides (Chen et al., 2010c) and attempts to correlate these features with structural and biophysical characterisations (see Chapter 7 and 8).

2.7 Concluding remarks

Hydrogelation is a phenomena that already has and could have an even greater repertoire of applications. The self-assembly basis for molecular hydrogels is also exploitable. There are many routes available to self-assembly and hydrogelation but there is a current interest in short peptidic aromatic conjugated LMWG. The mechanism through which the self-assembly of these systems can occur is variable but one includes the protonation mechanism whereby the peptide C-terminus is protonated to induce self-assembly.

The link between the pH at which gelation occurs and the pKa of the ionisable C-terminus is quite clear however there is no certain relationship between dipeptide molecular structure, assembly, fibril structure and material properties. The material properties of these systems are reproducible when assembly is triggered in the same manner (Adams et al., 2009) and so in theory the factors that determine material properties could be deconstructed. However the material property determinants are numerous and dynamically affect one another; including

fibril dimensions, degree of cross-linking, which are presumably dependent on molecular structure and kinetics of gel formation (Adams, 2011). It is unlikely that understanding any one of these parameters in isolation will hold the key to the link between LMWG and hydrogel material properties but to do so is an important step in this direction.

The current understanding of how the LMWG relates to the hydrogel formed is summarised scrupulously by van Esch (van Esch, 2009):

> This lack of insight into molecular gelation phenomena can be traced back to two different origins. First, there is a large gap between our design efforts and actual knowledge of the supramolecular structure of the gels (i.e., the molecular arrangement of gelator molecules within the gel fibers, the junction zones, and the fiber-liquid interface). As a consequence, the feedback loop between molecular design and experimental verification remains at the qualitative gelation-no gelation level, without leading to significant improvement of design and insights. This situation will change only when more detailed information on the molecular arrangement with the gel fibers becomes available. Limited knowledge on the supramolecular structure is not exclusive for molecular gels but reflects a general issue in supramolecular science because current experimental techniques remain inadequate to provide structural details at the supramolecular level...

The work contained in the chapters of this thesis pertaining to the investigations around LWMG aim to address this.

3 Techniques by which to study self-assembly and structure

3.1 Introduction

Structural information about self-assembled systems cannot readily be accessed through the standard structural biological suite of techniques, particularly in the case of amyloids, therefore a range of techniques must be employed to gather structural information. Often including a variety of biophysical and structural techniques across a wide resolution range and reporting on different aspects of structure, these techniques may include transmission electron microscopy (TEM), cryo-electron microscopy (cryo-EM), small angle X-ray scattering (SAXS), Fourier transform infrared spectroscopy (FTIR), circular dichroism (CD), linear dichroism (LD), nuclear magnetic resonance spectroscopy (NMR), single crystal X-ray diffraction (SC-XRD), powder X-ray diffraction (P-XRD) and X-ray fibre diffraction (XRFD). When combined, the data from these techniques is able to produce the most comprehensive models of the self-assembled system being described.

The following sections describe the theoretical background to the techniques used in this thesis.

3.2 Circular Dichroism

CD allows for the determination of the secondary structure of proteins in solution. As a solution technique it probes the native hydrated form of the proteins in question. Further, it can report information on dynamic changes in structure, data which is otherwise unobtainable by many other structural techniques.

3.2.1 Circularly polarised light and spectropolarimeter instrumentation

The production of circularly polarised light (CPL) may be achieved by the manipulation of linearly polarised light (LPL). Figure 3.1 illustrates this process. Briefly, a linearly polarised propagating waveform at 45° from vertical (a) may be combined with another linearly polarised waveform at 90° (b). Through constructive interference the resulting waveform will have a vertical plane of polarisation. However, if the two waves are adjusted out of phase by ¼ λ then the plane of the resulting wave rotates about the axis of the wave propagation (c), thus CPL. Left or right-handed CPL can be produced depending on which of the two waveforms are retarded. The production of CPL forms the basis of a spectropolarimetry.

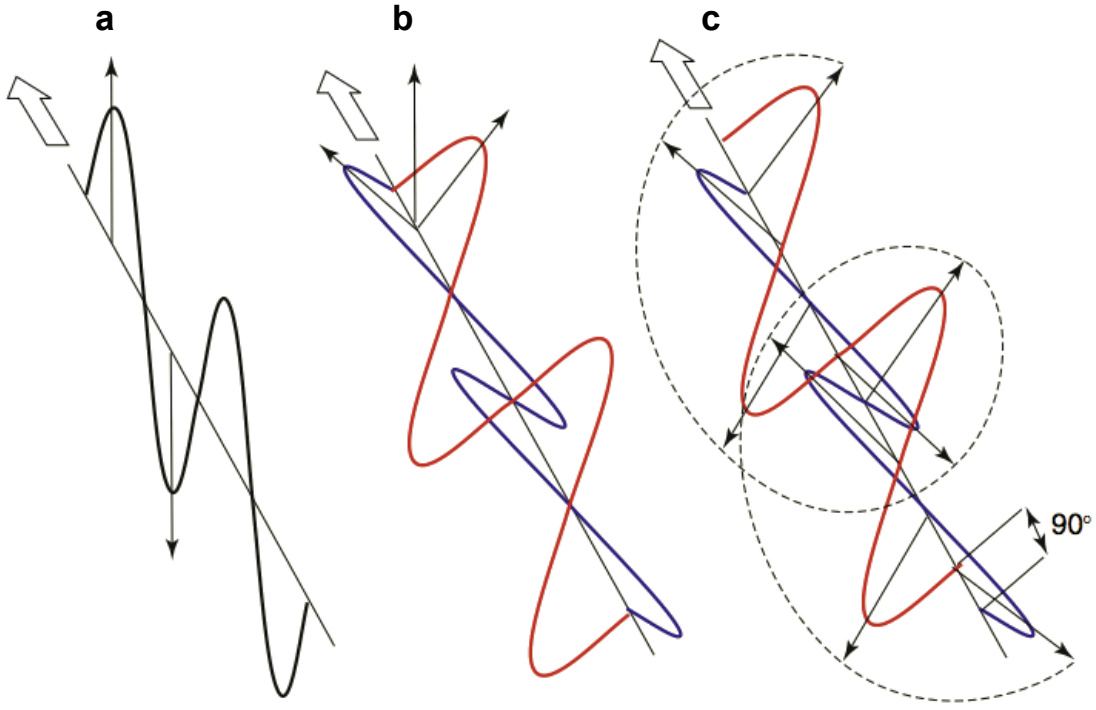


Figure 3.1: The production of circularly polarised light may be achieved by the manipulation of linearly polarised light. A linearly polarised propagating waveform (a) may be combined with another linearly polarised waveform at 90° (b); through constructive interference the resulting waveform will have a vertical plane of polarisation. If the two waves are adjusted out of phase by ¼ λ then the plane of the resulting wave rotates about the axis of propagation (c). Reproduced with permission from (Land, 2008).

Figure 3.2 shows the basic instrumentation of a spectropolarimeter designed to measure CD. Light is produced from a xenon arc lamp, which is passed through a series of optics to obtain monochromatic light; this is subsequently passed through a series of filters to create parallel and perpendicular LPL. The LPL is processed by a photoelastic modulator (PEM) oscillating at 50 kHz. The birefringent properties of this material allow it to retard one of the linearly polarised waveforms relative to the other. The piezoelectric properties of this material allow it to oscillate between states that alternately retard either the parallel or perpendicular LPL. The PEM as a result creates left and right-handed CPL at 50 kHz. The L- and R-CPL may interact with the sample, the transmitted light is digitised and amplified by a photo-multiplier tube (PMT) and subsequently detected. CD is measured by the difference in absorbance of either the left or right-handed forms of light (Equation 3.1), absorbance of either form causes the shape of the resulting light to deviate from a circle to an ellipse and so CD is traditionally measured in ellipticity or degrees, which is measured as positive or negative signals. Positive signals result from more L-CPL than R-CPL absorbed, negative signals from more R-CPL than L-CPL absorbed and zero signals arise from achiral molecules (Kelly et al., 2005).

$$CD = A_{L-CPL} - A_{R-CPL} \qquad \textit{Equation 3.1}$$

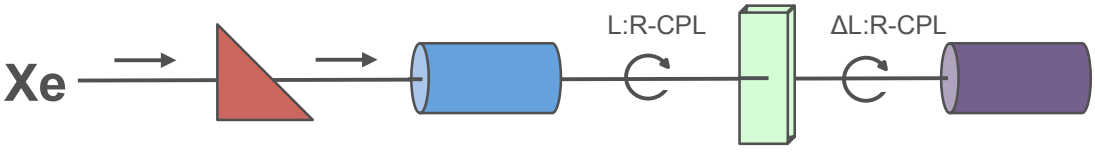


Figure 3.2: The optics of a spectropolarimeter. Chromatic light is emitted from a xenon source and filtered by a series of optics and a monochrometer (red) to create linearly polarised monochromatic light, this is passed through a photoelastic modulator (blue) to retard waves by ¼ λ creating circularly polarised light, the L:R-CPL interacts with the sample (green) and differential absorbance measured by a photomultiplier tube and detector (purple).

The differential absorption of L-CPL and R-CPL is a phenomenon exhibited by optically active chiral molecules; as such this is greatly beneficial to the structural investigation of proteins, which are chiral in nature.

3.2.2 Chiral amino acids and the amide bond

Chiral molecules are said to be optically active because the chromophores they contain will interact with/absorb CPL to differing degrees at various wavelengths depending on their enantiomeric nature. Naturally, all of the amino acids except glycine are optically active and as such, when CPL interacts with a polypeptide chain the absorbance of left or right-handed CPL will vary as a function of wavelength producing a characteristic spectrum. As shown in Figure 3.3 CD can distinguish between left and right-handed enantiomers of peptide chains through the noticeable inversion of the signs of the measured signals.

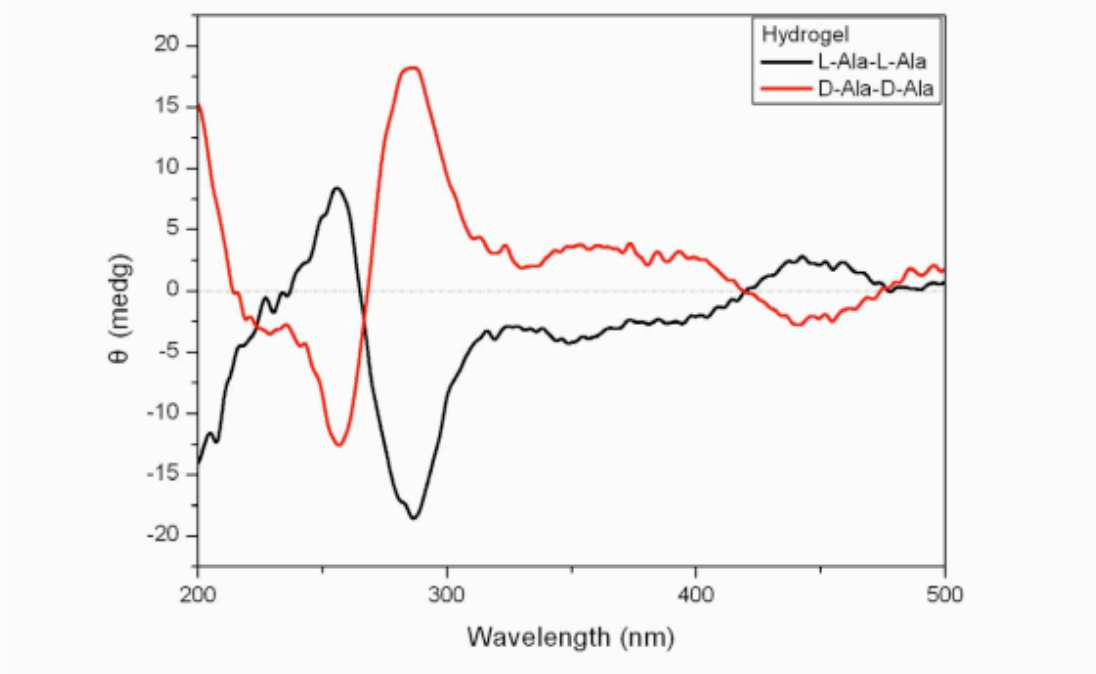


Figure 3.3: The CD spectrum from two enantiomers of the di-peptide spiropyran conjugated dialanine. Reproduced from (Qiu et al., 2009).

As photons of excitatory light interact with an electron orbital, the energy absorbed excites or causes the transition of electrons from low to high-energy orbitals, the redistribution of charge creates an electronic dipole moment (Rodger and Norden, 1997). The transitions that dominate a CD spectrum from a polypeptide arise from the peptide/amide bond. Figure 3.4 shows the amide bond (a) has excitable π-orbitals (b) and associated with the two lone electron pairs of the carbonyl oxygen two excitable *n*-orbitals (c). The amide bond absorbs between 190 – 230 nm where the π_o-π* transition occurs ~190 nm and the n-π* at ~220 nm (Fasman, 1996).

These transitions are sensitive to their environment and the conformation of the molecules they are found in. The n-π* transition is sensitive to environment where in apolar solvents can be at wavelengths as long as ~230 nm but in polar hydrogen bonding environments can be at wavelengths as short as ~210 nm. In contrast, the π_o-π* transition exhibits exciton coupling that is characteristic of particular secondary structural elements i.e. α-helices or ß-sheets. They have been well characterised theoretically and experimentally (Fasman, 1996) and as such their behaviour in proteins of different conformation are reasonably well understood.

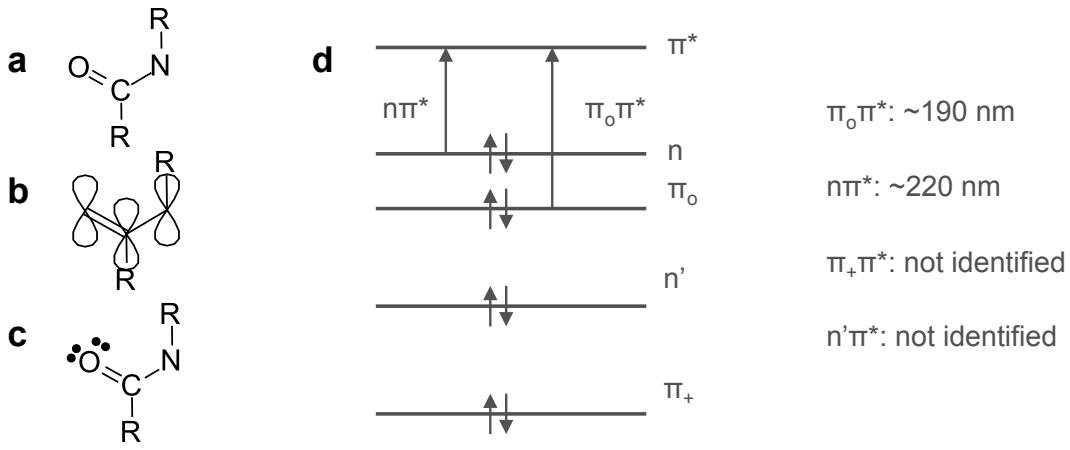


Figure 3.4: The electronic configuration of the amide bond (a) showing the three π-orbitals (b) and two lone pairs on the carbonyl oxygen (c). The electronic transitions in the far-UV region are shown where n and n' relate to the oxygen lone pair and π_+, π_o and π to the π orbitals (d) of the amide bond. Adapted from (Woody, 1996).*

3.2.3 Far-UV – Amide bond contributions

In the context of protein structure, the differential absorbance of L- and R-CPL through π_o-π^* and n-π^* transitions will depend on the environment those transitions occur in and thus the conformation of the protein, specifically the Φ and Ψ angles of the polypeptide chain. In its simplest use, CD can be used to identify secondary structure elements based on spectral shape. Figure 3.5a shows the typical spectra exhibited by the various secondary structural elements found in proteins. These characteristic spectra have typically been taken from model systems that adopt the appropriate secondary structures. The secondary structure of Poly-L-Lysine is sensitive to the conditions it is in and can be made to adopt either α-helices, ß-sheets or random coil and provides the basic spectral shape for protein secondary structure identification (Townend et al., 1966). Although broadly correct, it should not be overlooked that exceptions in the spectra exhibited by a protein in a particular conformation can occur.

The α-helical spectrum is well understood experimentally and theoretically, is characterised by a strong maxima at ~190 nm and minima doublet at ~210/222 nm. The π-π^* transition in α-helices displays exciton coupling into linear components with a positive component below 190 nm and a negative component at ~201 nm, the n-π^* transition has a single negative component at ~222 nm (Gratzer et al., 1961). The ß-sheet spectrum is less well standardised where it has been noted that the electronic transitions of this structure are quite sensitive to precise structure (Robert W, 1993 pg529). The ß-sheet poly-L-lysine is characterised by a single maxima at ~195 nm and minima at ~217 nm (Townend et al., 1966). The origins of these signals are more straightforward with importantly a π-π^* positive component at ~195 nm where exciton splitting is less evident and not interfering with the 190 – 250 nm region (Sreerama and Woody, 2004 pg318). The π-π^* transition is blue shifted in one to two ß-strand systems but red shifted in ß-sheet systems containing more ß-strands (Woody, 1996).

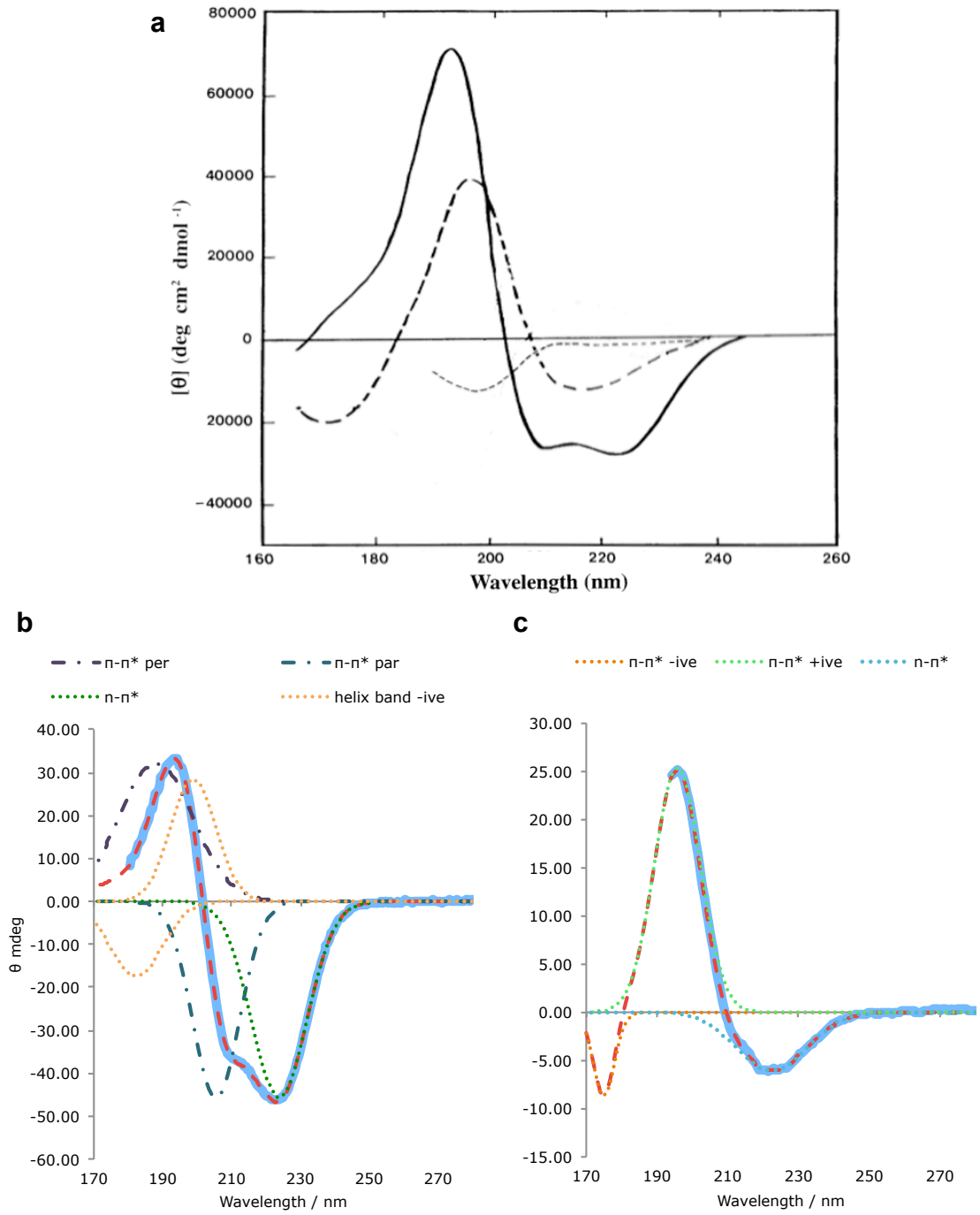


Figure 3.5: The typical CD spectra associated with various types of protein secondary structure, including α-helices (solid line), ß-sheets (long dashed line) and unstructured (dashed line) (a), reproduced with permission from (Kelly et al., 2005). CD spectra are a complex transition convolution of signals to form the resulting CD spectra of a α-helical or ß-sheet protein (b and c). Figure created by fitting three Gaussian peaks to a ß-sheet CD spectrum and minimising the resulting spectra by least squares as a reproduction from values in (Woody, 1996).

The n-π* transitions from ß-sheets are equally variable, for a ß-sheet forming fragment of $ß_2$-m (20-41) the n-π* transition is reported as short as 210 nm (Iwata et al., 2006). The n-π* in ß-sheets has a single negative component giving rise to a minima at ~215 nm. The transitions are shown in Figure 3.5b-c by deconvolution illustrating the level of complexity in the contribution of individual components to the resulting spectra measured.

3.2.4 Near-UV - Contributions from other chromophores

The aromatic amino acids tyrosine, tryptophan and phenylalanine as well as disulphide bonds, make contributions to the far and near-UV regions. Far-UV transitions are complicated by the presence of transitions from the amide backbone of polypeptides but in the near-UV the signals arising from aromatic residues can be well resolved and provide information about tertiary structure, protein folding and ligand binding (Woody, 2007). As with the amide bond, the transitions associated with aromatic chromophores are linear and aligned to specific axes, thus in a CD spectrum the resulting spectrum is a convolution of these bands. Table 3.1 summarises the transitions from phenylalanine, tyrosine and tryptophan.

	Assignment	λ_{max} / nm	ε_{max} / $M^{-1}cm^{-1}$	θ / degrees
Phe	Lb	260	200	90
	La	210	10000	0
	Bb	180	30000	90
	Ba	180	30000	0
Tyr	Lb	275	1400	90
	La	230	8800	0
	Ba	190	24000	90
	Bb	190	24000	0
Trp	Lb	280	3200	+41
	La	270	4500	-44
	Bb	220	35000	+11
	Ba	195	20000	-55

Table 3.1: The linear electron transitions in the aromatic sidechains phenylalanine, tyrosine and tryptophan. ε_{max} reflects the energy required to induce the transition and thus intensity of signal in spectroscopy. The transitions are assigned with the Platt notation for the excited states of benzene; angles of the transitions are given relative to the long-axis of the aromatic molecule axis. Values from (Woody, 2007).

The high-energy transitions arising from the B excitation bands are degenerate and relatively strong but overlapping with the stronger π-π* transitions of the amide backbone. The Lb and La excitation bands however are resolvable in some CD experiments, although their intensity is about two orders of magnitude less than signals in the far UV (Bulheller et al., 2007). The low ε_{max} values and also the fact that proteins have relatively low aromatic content compared to amide bond content means near UV aromatic transitions tend to be quite weak. The near UV region of CD spectra can give information about the chiral environment that aromatic residues occupy. The exciton-coupling phenomenon specifically can report on this. Where aromatic residues may stack and find themselves in a chiral environment, the excitation bands will split into positive and negative components. The two peaks from an exciton coupling, referred to as a couplet, can often be overlapping and so deconstructively interfere. However when resolved a right-handed arrangement will give rise to a negative short-wavelength and positive long-wavelength band and vice versa (Simonyi et al., 2003, Woody, 2007).

3.2.5 Exciton coupling geometry

Excitonic couplings are strongly dependent on the distance and relative orientations of the chromophore couplets (Mukamel et al., 1997). Coupling has been observed in porphyrin systems of up to 50 Å (Matile et al., 1996, Oancea et al., 2003). Strong coupling is defined as when the band splitting is greater than the width of the band from the monomer; cases of intermediate and weak coupling exist in the opposite cases (Kasha, 1963). Using the simple geometrical models of weak molecular exciton coupling proposed by Kasha *et al.* (Kasha, 1963) it may be possible to determine structural constraints on the geometries of exciton coupled chromophores from the stokes shifts in their corresponding absorbances. As shown in Figure 3.6, a monomeric chromophore (Figure 3.6a-i) may be closely associated with a neighbouring chromophore (Figure 3.6a-ii) but through weak interaction of the electronic orbitals result in splitting of the energy levels that the new exciton coupled structure can be excited to.

The model describes two polarised transitions that are parallel and also lie on the same plane, the parallel planar geometry. In this arrangement, variation is introduced by differences in the angle between the polarisation of one transition and the mid-point of the neighbouring coupled transition (θ). Two geometrical scenarios may first be considered, the end-to-end and face-to-face geometries (Figure 3.6a-ii and a-iv respectively). In the end-to-end arrangement (also known as J-aggregated, named after Jelley, one of the first investigators to classify these phenomena), where θ = 0 – 54.7°, the in-phase coupling of electronic orbitals results in a low-energy excitable state thus causing a red shift of the absorbance peak for the exciton relative to the monomer (Figure 3.6a-ii). The converse high-energy excitable state is true for the face-to-face arrangement (also known as H-aggregated, for hypsochromic shift) where θ = 90.0 – 54.7°, resulting in a blue shift of the exciton away from the monomer (Figure 3.6a-iv) (Kasha, 1963). In the parallel planar arrangement, all of the angles of θ between the ideal end-to-end (θ = 0°) and face-to-face (θ = 90 °) geometries result in lessening of the band splitting, thus varying the degree of red or blue shift. At an angle of 54.7° no band splitting and thus no shift would be exhibited (Figure 3.6a-iii) (Kasha et al., 1965).

Two coupled transitions may lie on the same plane but be non-parallel with regard to their respective polarisations, also described as an oblique geometry (Figure 3.6a-v). These arrangements are also described by the model proposed by Kasha *et al*, and result in both the low and high-energy excited states being permissible where both have components of in-phase coupling, causing observed band splitting of the original monomer excitation band (Figure 3.6a-i) (Kasha, 1963).

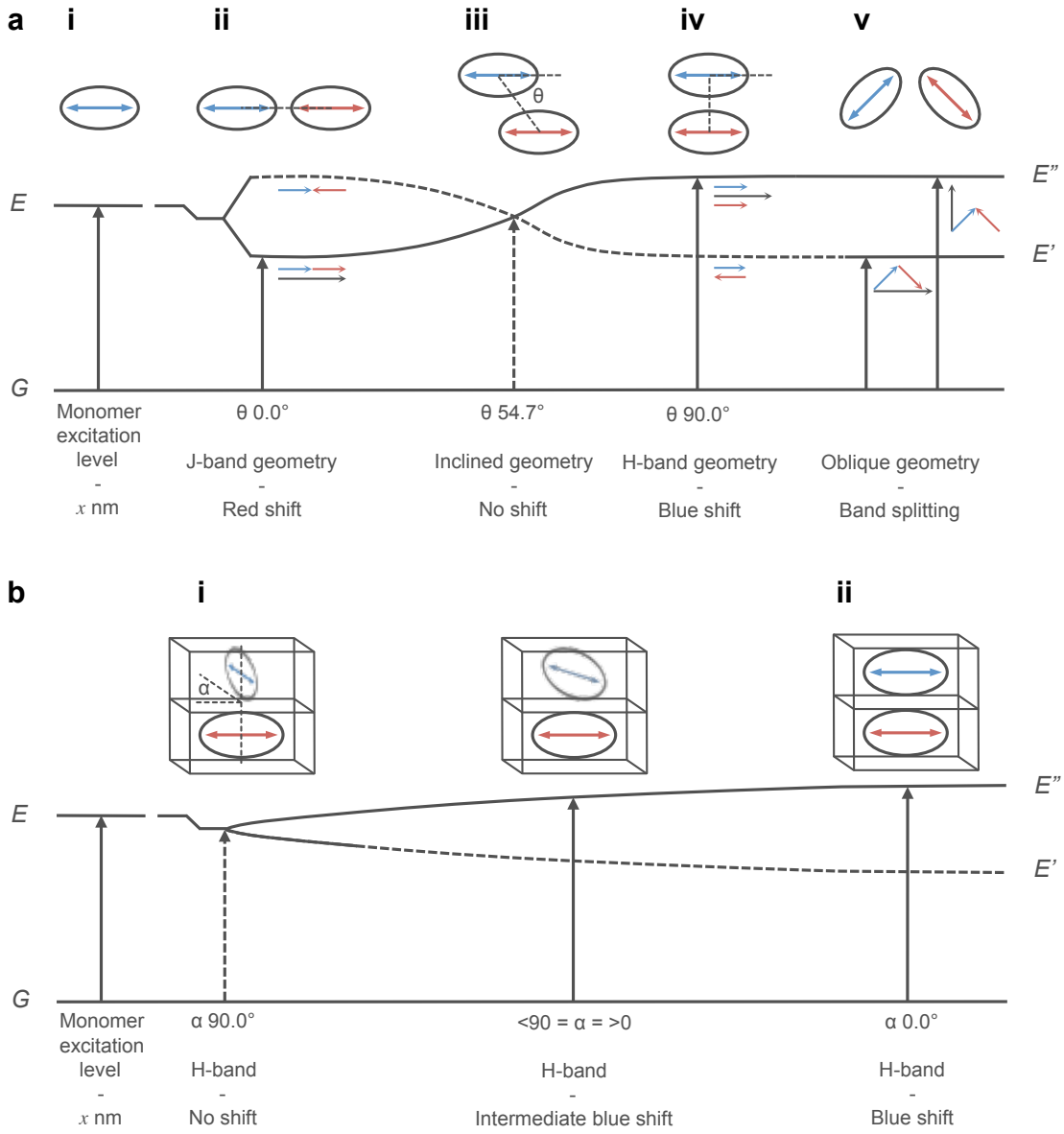


Figure 3.6: Exciton coupling theory and geometry. The parallel planar geometry of two polarised electronic transitions in an end-to-end arrangement results in a permissible low-energy coupling (a-ii), whereas the face-to-face arrangement results in a high-energy coupling (a-iv), non-parallel planar geometries result in band splitting (a-v). Further arrangements can be considered by the exciton coupling of non-planar geometries (b). Figure adapted from (Kasha, 1963, Kasha et al., 1965).

Further complexity becomes apparent when non-planar exciton couplets are considered, as shown in Figure 3.6b. One respective monomer may move through the angle α into a non-planar configuration decreasing the evidence of band splitting, where at the most non-planar angle α = 90° (Figure 3.6b-i) no exciton splitting is observed. Splitting increases as a function of increasing planarity to the case where α = 0° (Figure 3.6b-ii) and maximal blue shift is observed

(Kasha et al., 1965). The theoretical framework established for the dimer exciton condition was found to also be applicable to exciton coupling in the polymeric condition and so could be applied to understand the geometries of chromophores in polymeric assemblies (Kasha, 1963).

3.2.6 Exciton coupling helicity

As described in Section 3.2.5, the geometry of two neighbouring exciton coupled chromophores has an affect on the resulting absorbance profile. In the case where exciton coupled transitions are in the oblique geometry, both the high and low-energy transitions are permissible and band splitting is observed around the uncoupled absorbance band of the monomer. CD is uniquely able to identify the helicity of such exciton coupled geometries through the observation of Cotton effects, as shown in Figure 3.7. Cotton effects (as first described by Cotton) are characterised by the rapid change in the sign of a CD signal where the zero point of this change corresponds to the absorbance peak of the uncoupled monomer (Fasman, 1996). Two coupled transitions in a left-handed helical arrangement will exhibit a negative Cotton effect where the longer wavelength transition will be negative (Figure 3.7a). In contrast a right-handed helical arrangement of coupled transitions will have a positive Cotton effect (Figure 3.7b) (Harada and Nakanishi, 1972).

Where band splitting due to exciton coupling may be small, in a normal absorbance profile this phenomenon may not be detected (Figure 3.7c-d). However due to the Cotton effect, in the general case, exciton coupled CD provides an unambiguous method for the determination of the helical configuration of molecules (Rodger and Norden, 1997).

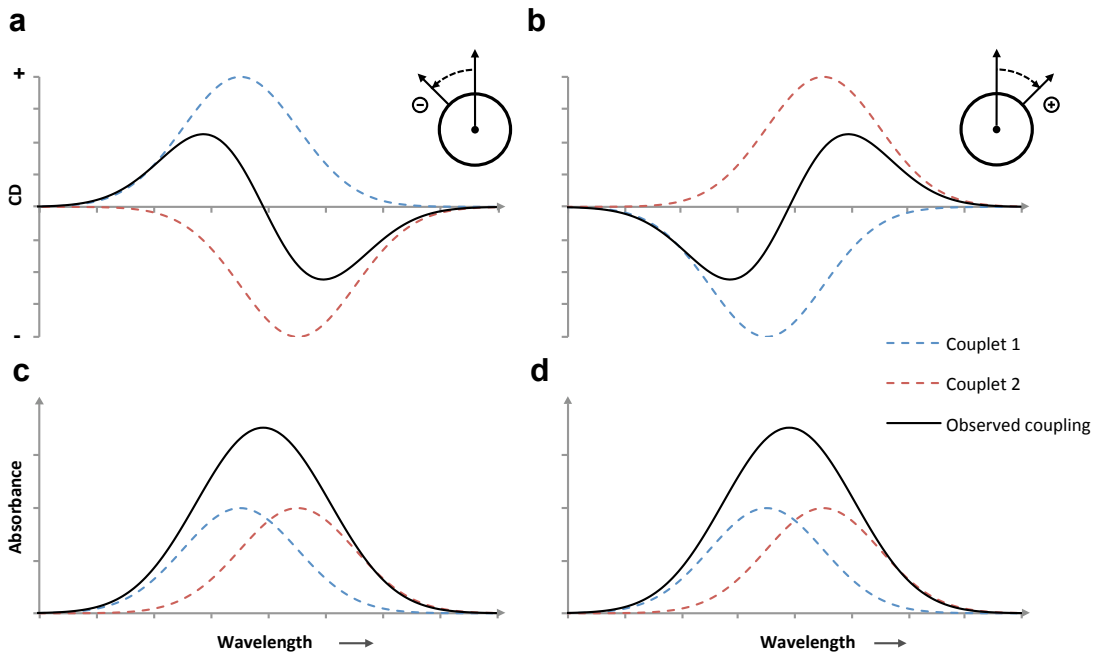


Figure 3.7: An illustration of the exciton CD exhibited by differing helical arrangements of transitions. An anti-clockwise screw or left-handed helix will give rise to a negative Cotton effect (a) and a clockwise screw or right-handed helix will give rise to a positive Cotton effect (b). In the corresponding absorbance profiles the same band splitting may be obscured by constructive interference (c-d). Peaks are produced as a Gaussian function of wavelength.

3.2.7 Experimental considerations – in general and with self-assemblies

A number of factors must be considered to collect quality CD spectra. These factors can all be easily understood when considering CD is essentially a type of absorbance spectroscopy and thus data quality relates directly to balancing absorbance intensity against sample scattering and artefacts. The quality of CD data can be judged by the signal to noise ratio in the CD spectrum and the high tension voltage (HT[V]) profile. Figure 3.8 shows a raw measurement of an ß-sheet protein structure characterised by CD. The noise is shown as an insert and varies by approximately 2,000 deg cm^2 $dmol^{-1}$ with baseline shifting over approximately 6,000 deg cm^2 $dmol^{-1}$, the signals are ~100 orders of magnitude over this and would be considered of high enough quality. To improve data quality, two experimental parameters may be adjusted: the sample concentration or cell path length.

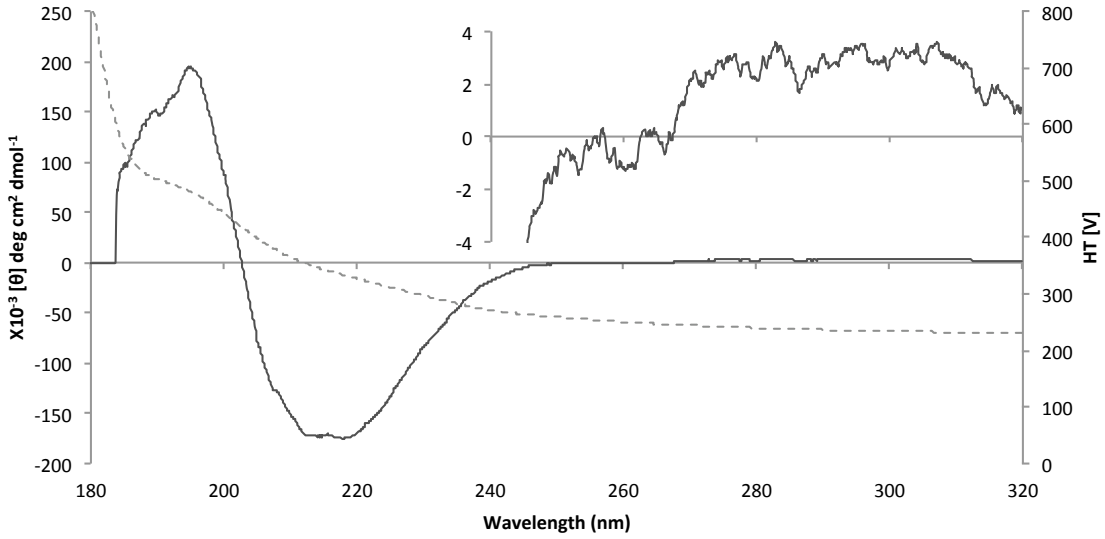


Figure 3.8: A high quality CD spectra from a protein in a ß-sheet conformation. CD is shown as a solid line, the HT[V] as a dotted line. The inset shows variation due to noise.

For far UV experiments quartz cuvettes of between 0.01 – 1 mm are used, as path length is increased the amount of material interacting with the incidence light is increased thus increasing signal intensity and also signal to noise ratio however the HT[V] profile will increase. The HT[V] profile indicates at a particular wavelength the voltage applied across the PMT and thus how much this piece of apparatus is having to amplify the signals/photons transmitted through the sample. A high HT[V] may indicate either scattering effects or strong absorbance but ultimately is a reflection of fewer photons reaching the PMT detector because of either sample scattering or absorbance effects. In each case above a certain threshold, the PMT will have to amplify such small signals that any standard deviation or error in the number of photons reaching the PMT will be amplified as noise. The following experimental considerations relate directly to these described effects.

Buffer choice is important where carboxylic acid groups, chloride ions, imidazole, HEPES, MOPS, MES and PIPES all generally absorb strongly below 200 nm (Kelly et al., 2005) but high concentration complex buffer backgrounds will also have a high baseline scattering profile contributing to poor quality spectra collection. The scattering profile will also be affected by the

turbidity of sample in question and thus factors including solubility and size of solute influence the scattering profile. Both of these effects are directly proportional to sample concentration and fibrillar self-assemblies are well known to be insoluble and strongly scattering.

A special case to be considered with the characterisation of self-assemblies including amyloid with CD is the influence of anisotropy. The characterisation and interpretation of spectra from CD is based on the assumption that the molecules interacting with the incident light are isotropic; that is randomly distributed at all possible orientations in the solvent. The isotropic nature of the sample results in a spectrum that represents an average of all of the linear transitions exhibited by chromophores. In the juxtaposed scenario, if in an anisotropic sample the linearly polarised transitions align to an axis then the spectra from the sample will be directionally dependent, greatly affecting the interpretation of measurements made. Fibrillar self-assemblies can have a propensity to auto align or with the use of demountable cuvettes align by shear flow when loading the sample. In a CD experiment the absence of artefacts from linearly polarised transitions should be verified, however if the alignment of these polarised transitions can be controlled then this kind of analysis may reveal useful directional information about a sample.

3.2.8 Circular dichroism analysis

Due to the reasonable experimental examples and theoretical understanding of the origins of peaks in a CD spectrum one can usually visually recognise α-helical vs ß-sheet vs random coil signatures. CD can be a very useful comparative analytical tool where spectra may simply be overlaid to check for folded protein or change in structure upon change of environment or mutagenesis. Change in structure is often monitored at single wavelengths where α-helical content may be comparatively measured at ~222 nm and ß-sheet content at ~220 nm. Though this method has been successfully used throughout the protein folding community, it is done so

with disregard to overall spectral shape changes which in some circumstances may overlook useful structural information (Greenfield, 1996).

More thorough attempts at quantitative secondary structure content analysis from CD spectra began with mathematical calculation from n-π* signal intensities (Woody, 1996). These methods should in principle be very precise at determining secondary structure content but become highly complicated if dealing with proteins with mixed secondary or dynamic structures. More recent methods are based on an empirical comparison of known reference spectra to the experimental spectra in an attempt to estimate each kind of secondary structure. Dichroweb (URL: http://dichroweb.cryst.bbk.ac.uk/html/home.shtml) is a web-based server for the secondary structure content analysis of CD spectra (Whitmore and Wallace, 2004). It can use a number of algorithms and reference spectra for analysis but using the CDSSTR analysis algorithm and the SP175 reference set (Lees et al., 2006) conceptually the analysis can be described as follows; a model spectra is created by combining spectra from proteins of known secondary structure and comparing the model to the data to be analysed. In an iterative manner the model spectrum secondary structure content is adjusted based on variable contribution from each reference spectra until the model data has a high correlation of fit to the experimental data as measured by the normalised root-mean-square deviation (NRMSD) value (Mao et al., 1982). By calculating the contribution of each experimental reference spectra to the best fit model spectra it is possible to estimate secondary structure content.

3.3 Linear Dichroism

3.3.1 Theoretical background

As stated, the signals of a CD spectrum are a complex convolution of unique linearly polarised absorbance bands, as shown in Figure 3.9 (see also Table 3). In CD these are averaged due to isotropy in the samples whereas in an LD experiment an anisotropic sample is prepared so that linear transitions are aligned to specified axes. All electronic transitions are polarised and when

they interact with LPL with the same polarisation as the transition absorbance is maximal (Rodger, 1993). In this case information about the relative orientations of chromophores can be obtained (Dafforn et al., 2004, Dafforn and Rodger, 2004). Accordingly, LD is measured by the difference in absorption of LPL parallel ($A_{//}$) and perpendicular ($A_{\perp}$) to an orientation axis as generally shown in Equation 3.2.

$$LD = A_{\parallel} - A_{\perp} \qquad \textit{Equation 3.2}$$

The orientation axis is by convention, in a couette flow set up, the horizontal axis and so a linear transition aligned horizontally will have an LD signal of > 0 but one aligned vertically will have an LD signal of < 0. The intensity of the LD signal will then decrease as the linearly polarised transition deviates from the horizontal or vertical plane and thus LD signal intensity can be used to calculate the orientation of the chromophore relative to the orientation axis. The relationship between the angle of the electronic transition and the LD signal intensity is shown in Equation 3.3, where the reduced LD signal (LD^r) is calculated by normalising the measured LD signal (LD) to the contribution of isotropic absorbance (A_{iso}). LD^r may then be related to the angle between the orientation axis and the measured transition polarisation (α) if the degree of orientation (S) is considered.

$$\frac{LD}{A_{iso}} = LD^r = \frac{3}{2} S(3\cos^2\alpha - 1) \qquad \textit{Equation 3.3}$$

3.3.2 Linearly polarised electronic transitions

Linearly polarised electronic transitions have been discussed in Section 3.2 but are summarised in Figure 3.9. These arise from the same excitation of π and *n* electronic orbitals. If the polarisation of one transition is known then it may be possible, using Equation 3.3, to calculate the polarisation angle of other transitions, thus revealing orientation information about chromophores.

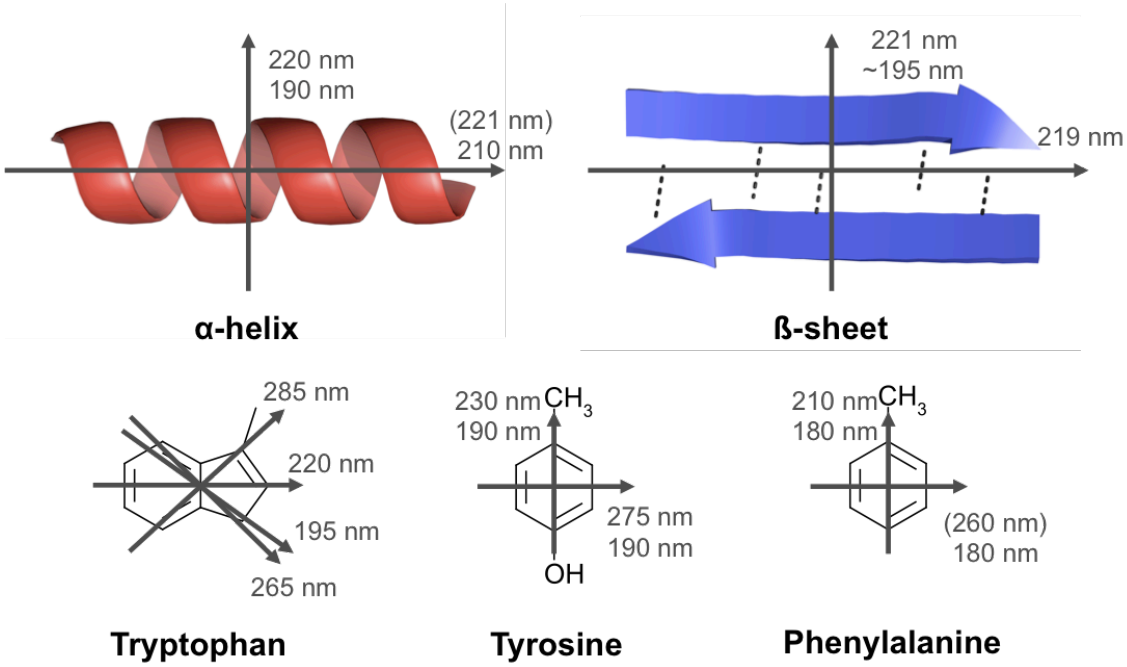


Figure 3.9: The linear electronic transitions from the common secondary structural elements and aromatic chromophores. Weak transitions are indicated by parenthesis. Adapted from (Hicks et al., 2010).

3.3.3 Sample alignment

The key to the successful use of this technique is the ability to align chromophores so that Equation 3.3 can be reliably used. In the case of fibrillar self-assemblies, this can be achieved by taking advantage of the behaviour of fibrils under flow. The use of couette flow cells has been developed for this purpose (Dafforn et al., 2004). Figure 3.10 shows the experimental set up using couette flow to align a fibrillar sample. This consists of a quartz cylindrical cell with an outer jacket and inner sample rod creating a low volume cell with specified pathlength. Rotation of the outer jacket establishes couette flow forces that any linear molecule should align with, by this manner fibrils may be aligned to the orientation axis. Fibrils may also align by shear forces on rod insertion during sample insertion.

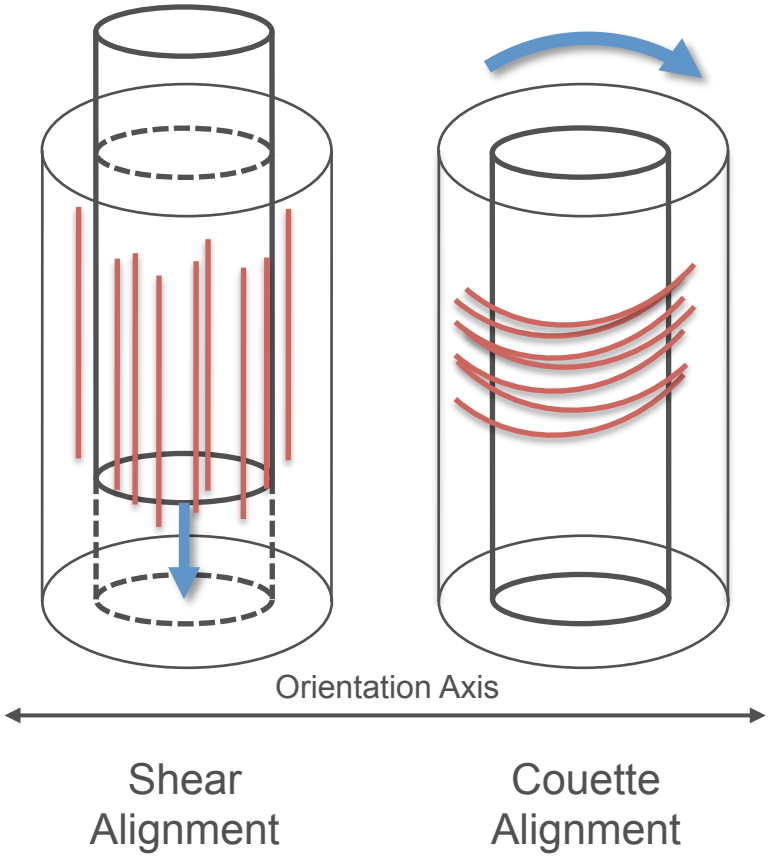


Figure 3.10: The use of couette flow to align fibrillar assemblies. The sample is housed into a low volume quartz cylindrical cell housed between the outer jacket and inner quartz sample road. Fibrils may be aligned by shear alignment upon loading the sample but upon rotating the sample rod couette flow is established aligning fibrillar material to the orientation axis.

3.3.4 Linear dichroism of self-assembled proteins

Linear dichroism in the UV/Vis region and the theory underlying the technique has long been established (Hofrichter and Eaton, 1976) but has recently experienced a resurgence of new interest due to its usefulness in characterising anisotropic molecules (Dafforn and Rodger, 2004). Amyloid-ß (Aß) in its fibrillar form was shown to give a positive LD signal at ~200 nm, where this signal arises from the π-π* transition polarised perpendicular to the ß-strand direction this data is consistent with the cross-ß model of amyloid (Dafforn et al., 2004). Other self-assembled fibrillar systems have been successfully characterised using UV-LD including the amyloidogenic short peptide GNNQQNY (Marshall et al., 2010), actin, Aß(1-42) and α_1-antitrypsin (Dafforn et al., 2004), glucagon amyloid fibrils (Andersen et al., 2010), $ß_2$-m (Adachi et al., 2007), α-helical forming coiled-coil self-assembled fibres (SAFs) (Bulheller et al.,

2009), Tropomyosin (Pandya et al., 2000), FtsZ (Bulheller et al., 2009) and collagen (Mandel and Holzwarth, 1973 12-pg655). It is noteworthy that the use of LD is not just confined to the UV/Vis region for the characterisation of such assemblies (Hiramatsu et al., 2004).

3.4 Transmission electron microscopy

3.4.1 Visualisation and Resolution

Transmission electron microscopy takes advantage of the higher resolution granted by using electrons as the scattered particle in a microscopy set up. An electron microscope operated at 100 kV with a wavelength of 3.9 pm can potentially resolve objects as close as 2.4 Å. However, this is limited by the magnetic optics of the microscope and the method of visualisation of samples for inspection. Samples in transmission electron microscopy are typically visualised using a heavy metal negative stain such as uranyl acetate. Figure 3.11 shows an electron micrograph one would expect to obtain from a fibrillar sample of amyloid-ß (Aß). Negative stain electron microscopy is a relatively rapid method for visualising samples but it must be taken into consideration that what is observed is not the biological sample itself but the lack of detection of transmitted electrons due to the strong scattering from the heavy metal stain (Saibil, 2007). Care should be taken in that TEM can overestimate the diameter of certain structures if they are subject to flattening on the grid surface, this has been observed by a comparison of atomic force microscopy (AFM), cryo-EM and TEM negative stain measurements (Xu et al., 2010). Figure 3.11c illustrates this effect, in the case of fibrillar samples negative stain will almost certainly adsorb to the grid at the edges of material (ci) and may penetrate finer structural features (cii & iii) but the absence of these staining features does not necessarily mean these kinds of features are not present. The sample preparation is an important factor to bear in mind, the dehydration process may distort structures and the staining conditions are generally acidic. Despite this the method is considered an effective way of obtaining shape, symmetry and structural arrangement information for a sample (Saibil, 2007).

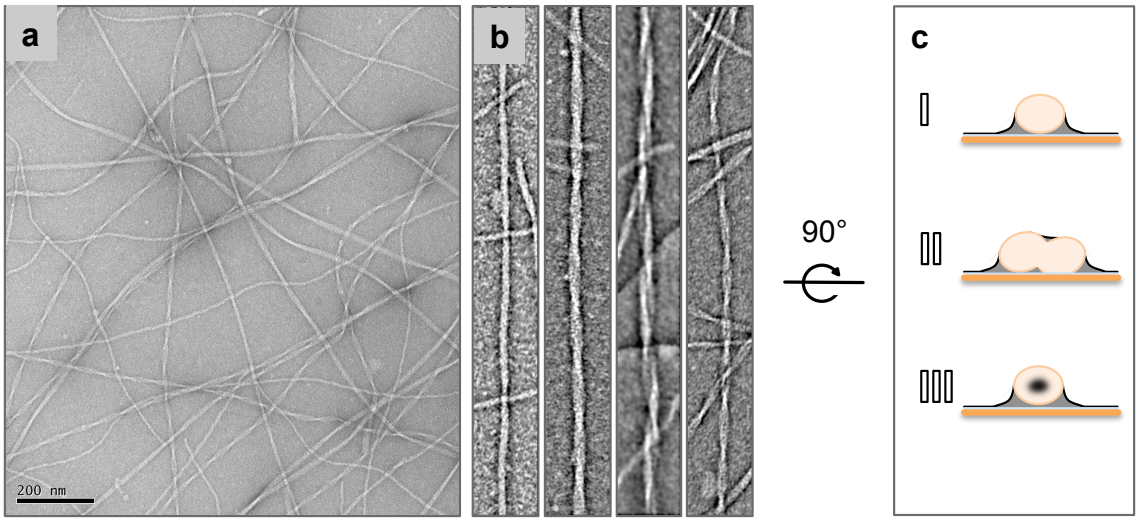


Figure 3.11: A typical negative stain transmission electron micrograph obtained for amyloid-ß at low magnification (a). High magnification micrographs often reveal the variable morphologies of amyloid fibrils (b) through stain penetration of the filamentous structure of the assemblies. Negative staining scenarios (c) could include single filament (ci), paired twisted filaments (cii) and nanotubular filaments (ciii). Figure reproduced from (Morris and Serpell, in press).

3.5 X-ray diffraction

3.5.1 X-ray diffraction theory

X-rays typically have energies ranging from 0.*120 - 120 keV* (1.92 x 10^{-17} - 1.92 x 10^{-14} J) and thus potential wavelengths of 10 – 0.01 nm as given by Equation 3.4 where wavelength (λ) is proportional to the speed of light (c = 3 x 10^{8} ms^{-1}), Planck's constant (h = 6.63 x 10^{-34} m^2kgs^{-1}) and energy (E). The X-rays used in a modern diffractometer may be produced by a rotating Cu Kα anode X-ray source where the wavelength of the X-rays are 1.5419 Å.

$$\lambda = ch / E \qquad \textit{Equation 3.4}$$

The wavelength of X-rays produced at this energy are roughly equivalent to the distance of interatomic bonding and so X-rays interacting with molecules undergo diffraction. It is useful to consider the diffraction pattern observed from the classical double slit experiment as shown in Figure 3.12a.

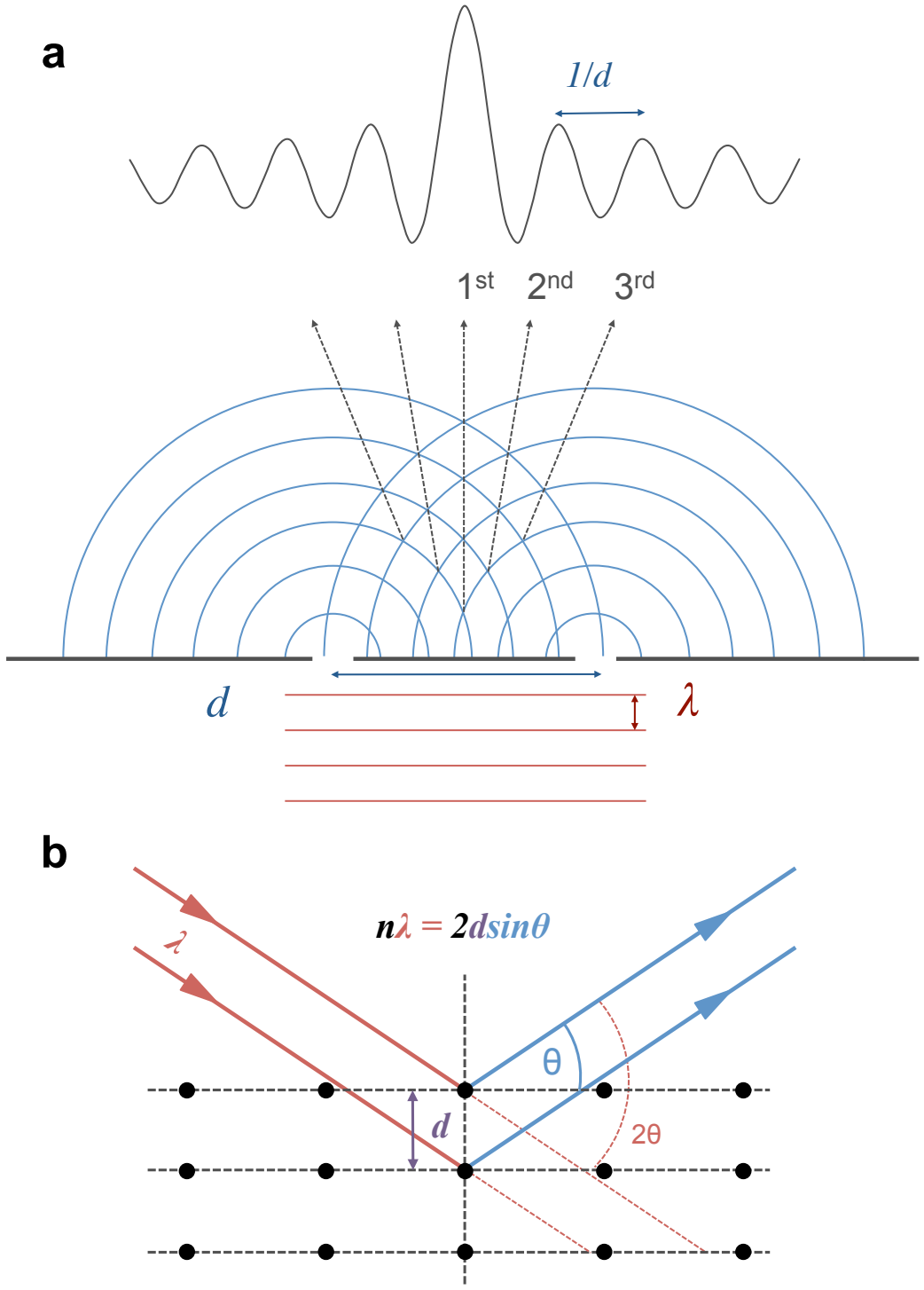


Figure 3.12: The diffraction of X-rays illustrated by the classic double slit experiment (a) and the interaction and diffraction of X-rays with atoms of a repeating lattice where constructively interfering waves satisfy Bragg's Law (b).

A waveform that encounters a double slit will diffract, the resulting diffracted waveforms will interfere constructively and destructively to produce a one-dimensional diffraction pattern. The intensity of the diffracted peaks is governed by a zero order Bessel function and can be rationalised whereby the 1st order reflection arises from the most number of constructive interferences and high order reflections from fewer interactions. The distance between the peaks is inversely proportional to the distance between the double-slit. A crystal lattice of atoms can be compared to diffraction slit experiments. Figure 3.12b illustrates the interaction with and subsequent diffraction of X-rays from the electrons of atoms in a crystalline lattice. X-rays diffracted from repetitively spaced atoms will constructively interfere and reinforce one another to create a diffraction signal called a Bragg peak. The diffraction signals or Bragg peaks angular position relative to the incident X-ray waveform is related to wavelength by Bragg's law as shown in Equation 3.5, where the lattice spacing (d) and angle between the incident beam and plane of scattering atoms (θ) is proportional to the order of the Bragg reflection (n) and the wavelength of the incident X-rays (λ).

$$2d\sin\theta = n\lambda \qquad \textit{Equation 3.5}$$

Thus any system with repetitive lattice spacing can potentially diffract X-rays such that interatomic spacing information might be obtained. Various X-ray diffraction techniques have been developed whose use is mainly dependent on the type of sample being investigated. These techniques range from single crystal X-ray diffraction and powder X-ray diffraction to fibre diffraction. X-ray fibre diffraction is essentially a technique that lies somewhere between powder X-ray diffraction and single crystal X-ray diffraction. All three techniques give information about the repetitive spacings of atoms in three-dimensional space whereby the positions of diffracted signals are directly related to interatomic distances. Where these techniques differ is in the degree of order in the system being studied.

3.5.2 X-ray diffraction techniques

In a perfectly ordered system, for example in the crystalline state, all atoms can be arranged in a lattice and are perfectly repeating. SC-XRD aims to crystallise a protein into this phase such that the protein in question is found arranged into a repeating lattice consisting of unit cells. Each unit cell is a perfect copy of every other unit cell in the crystal lattice and the unit cell dimensions can be used such that an atom in space is perfectly related to its identical atom in a neighbouring unit cell by translation through the lattice dimensions (Blow, 2002). Figure 3.13 shows an example of an orthorhombic (a) and monoclinic (b) unit cell. Unit cells often contain multiple copies of the crystallised molecule; the single principle molecule of the unit cell is called the asymmetric unit. Two further parameters can be used to describe crystal lattices: the space or symmetry group. Using these parameters it is possible, from the asymmetric unit, to create every other molecule in the unit cell based on translation and symmetry operations defined by the unit cell space group (Ladd and Palmer, 2003). Diffraction from the basic lattice will produce a two-dimensional projection of the three-dimensional lattice parameters whilst any structures inside the lattice will modulate the intensities of the diffraction spots. If the crystal is tested from all possible angles it is possible to integrate the diffraction data and produce of three-dimensional model of the electron density within the repeating lattice (Blow, 2002).

The contrasting scenario is found in the case of X-ray powder diffraction whereby diffraction is not carried out from a single crystal but a collection of microcrystals at random orientations. P-XRD gives interatomic distance information that is entirely radially averaged because the microcrystals within the sample are at random orientations with respect to one another. In this case the two-dimensional projection of the three-dimensional lattice will be the same however the sample is orientated. Interatomic distance information is contained in the pattern but must be deconvoluted to assign distances to particular unit cell axes.

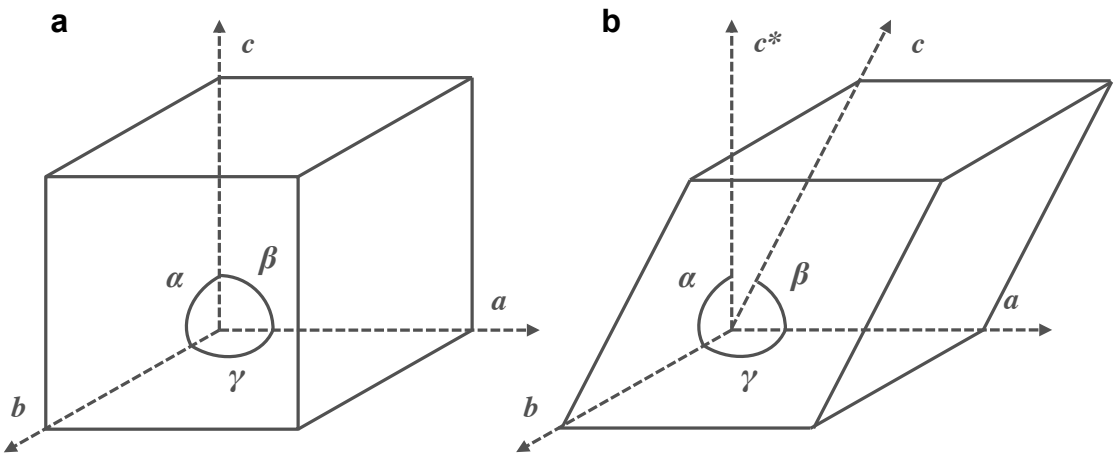


*Figure 3.13: The conventional crystallographic dimensions and angles of an orthorhombic unit cell where α=β=γ=90 (a) and a monoclinic cell where α=γ=90 β≠90 (b). In the monoclinic cell the reciprocal axis indicated by * remains perpendicular to a and b despite the change in the c dimensions, the same is true of the other dimensions with non-90 unit cell angles.*

3.5.3 X-ray fibre diffraction

Diffraction from a helix

XRFD concerns the diffraction of X-rays from repetitively structured fibrils. XRFD lies somewhere between SC-XRD and P-XRD techniques where two levels of disorder are apparent in a fibril sample: crystallinity of the fibre (crystallites) and the degree of order between crystallites. Self-assembled fibres are typically inherently insoluble, large and heterogeneous and hence are notoriously difficult to crystallise. However due to their self-assembled nature they have a crystalline order along their fibre axis, though not laterally perpendicular to the fibre axis. The degree of this crystallinity is variable and sample dependent. If a collection of fibres can be aligned, then a lattice of imperfect crystallinity can be created such that the sample may diffract X-rays. Figure 3.14 shows the idealised diffraction one would expect from DNA. A useful example in understanding the theory of fibre diffraction, a sample of DNA fibres may be aligned to a degree where they diffract X-rays (a-b). The resulting pattern will display a helical cross pattern that contains two main features relating to the repetitive atomic spacings in the DNA helix (c). The major reflections on the horizontal axis relating to spacings on the corresponding axis of the sample are reciprocally related to the distance between bases and the minor reflections away from the horizontal axis are reciprocally related to the distance between

each turn of the DNA helix. It is important to note that DNA has a relatively high crystallinity along its fibre axis and may be aligned to a relatively high crystalline sample for diffraction analysis. In the case that the DNA fibres were disorientated from one another (d) then the pattern would become much more challenging to interpret (e).

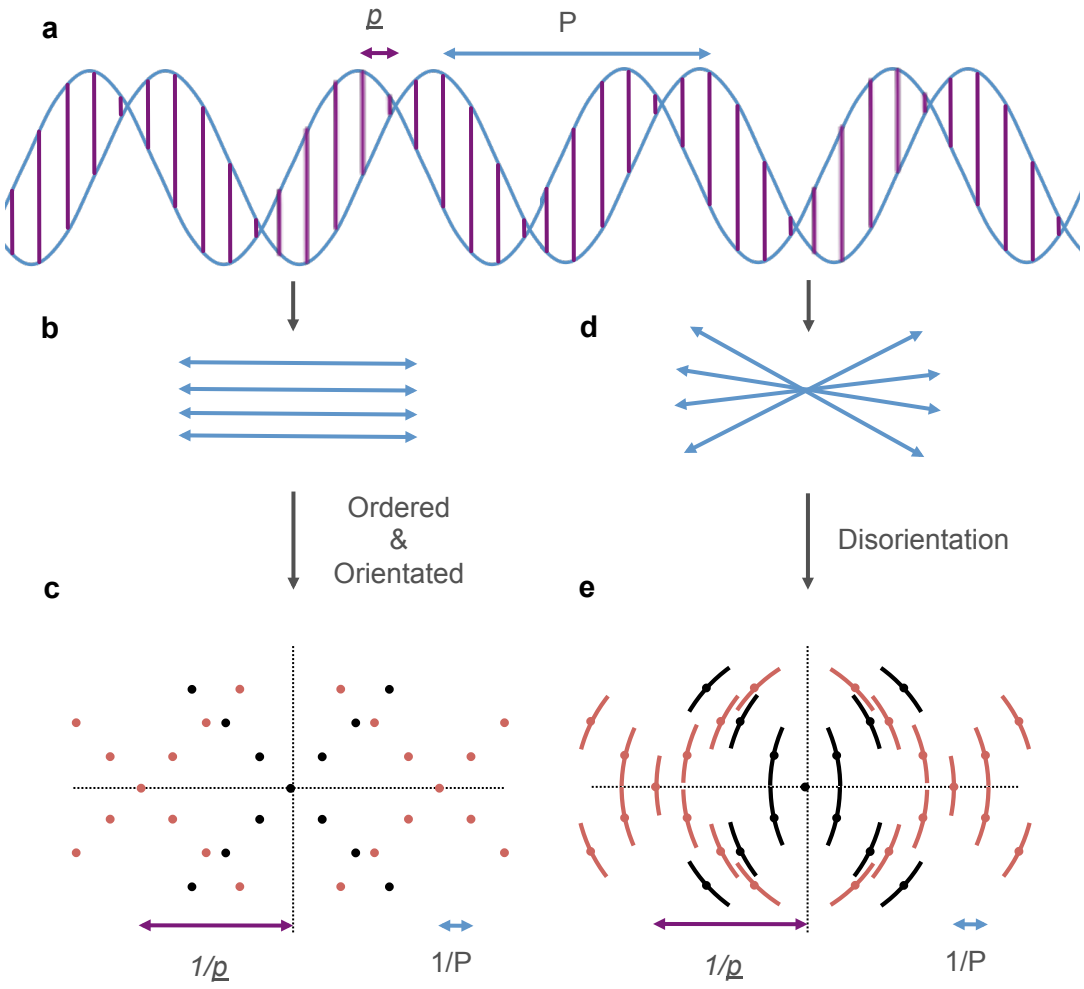


Figure 3.14: Idealised fibre diffraction from DNA. A double helix of DNA can be considered to be a discontinuous helix with two parameters; the pitch of the helix, P and the distance between subunits of the helix, p (a). In an ordered and orientated collection of DNA crystallites (b) the resulting fibre diffraction pattern exhibit a 'helix cross' (c). The separation of signals lying on the meridian is reciprocally related to P and the separation of signals lying off the meridian is reciprocally related to p. In the disorientated form (d) these signals become radially averaged (e) greatly increasing the complexity in layer line analysis.

Fibre alignment, disorder and crystallinity

Fibres may be aligned by a number of methods but all essentially involve the drying of a high concentration sample of amyloid fibrils. It is useful to consider a fibril sample to be made up of a collection of crystallites; each alignment method aims to align those crystallites such that their crystallographic axes are aligned in a known way. Figure 3.15 illustrates the various alignment methods frequently used in the preparation of samples for fibre diffraction experiments (Makin and Serpell, 2005b, Morris and Serpell, 2012). Figure 3.15a shows the most common and simplest texture to understand, here by dehydrating a sample of amyloid fibrils an attempt is made to align all fibre long axes such that they are parallel to produce a fibrous texture. In the case of Figure 3.15a all fibre axes are parallel and in Figure 3.15b-c (mat and disc textures) all fibre axes are aligned to the same plane, revealing information about structure but also potentially reporting on preferential fibre alignment.

The arrangement of crystallites with respect to one another is described as the sample texture (Alexander, 1969 pg 21, Makin et al., 2006). The fibrous texture of a fibre alignment is best illustrated in Figure 3.15d and the texture of the so-called mat or disc (analogous to a film texture) in Figure 3.15e. The degree of crystallinity, or perhaps better regarded as degree of disorder, has a direct effect on the quality of, and thus amount of information contained in, the diffraction pattern obtained. The crystallinity of a fibril alignment may be described as para-crystalline or disordered (Uri, 2008 pg 568). The para-crystalline state describes a collection of orientated crystallites that are randomly rotated about their long axis with respect to the fibre axis. Disordered fibre samples have little or no alignment in their crystallites on any axis and produce patterns similar to powder diffraction patterns. In each case, the coherence length describes the average length of these molecular segments/crystallites and deviation from parallel to the alignment axis is termed disorientation.

In the disorientated scenario, the likely reason for poor alignment and a disordered texture is due to a low coherence length and/or disorder within the crystallites themselves. P-XRD is often

able to provide more structural information for simple small molecule systems because the crystallites, even though disorientated, have a high crystalline order individually.

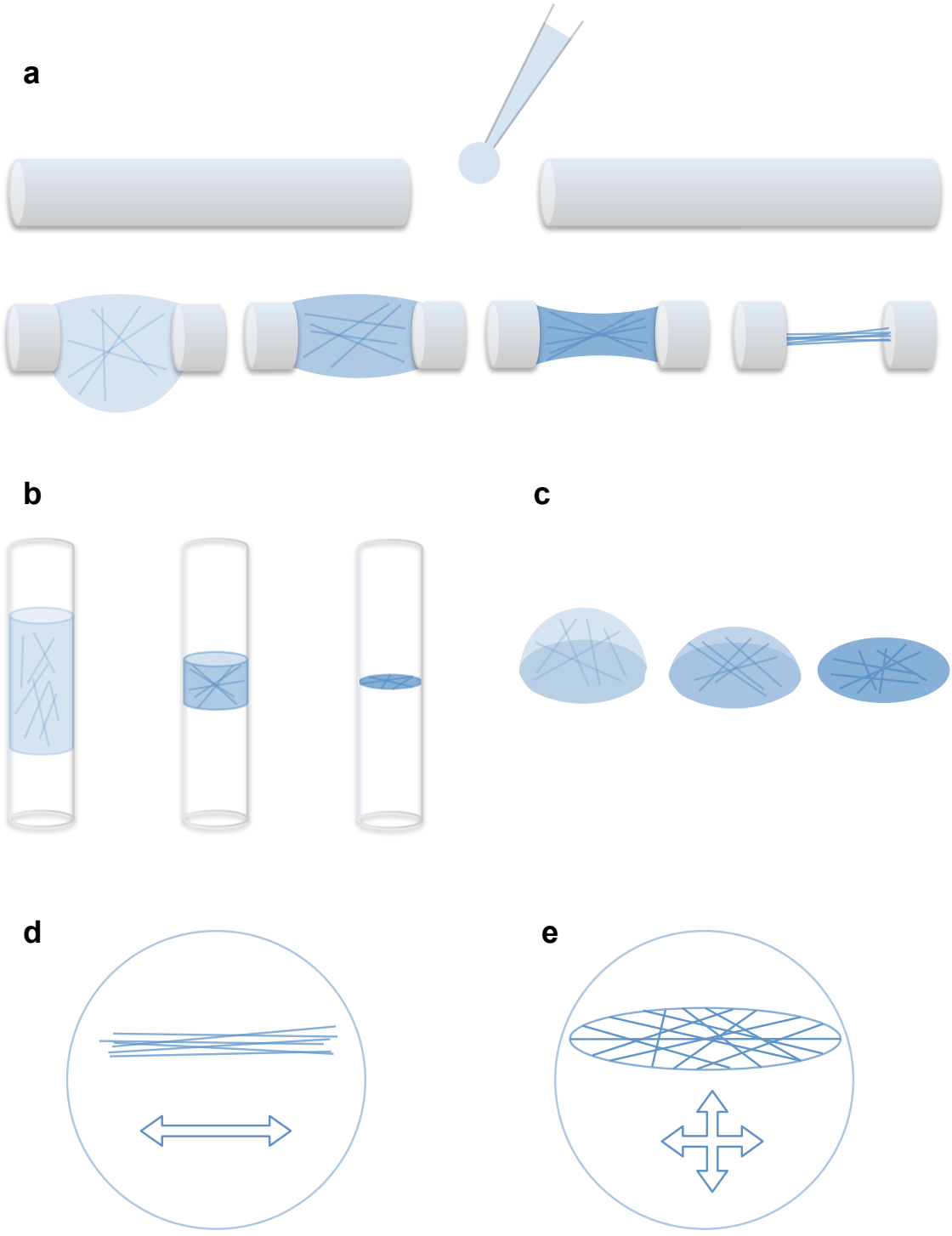


Figure 3.15: Methods of aligning fibrils for X-ray fibre diffraction may produce alignments with a fibrous texture (a), disc (b) or mat (c) texture. In the fibrous texture all fibre axes are parallel (d) and in the disc and mat texture the fibre axes are aligned to a single plane (e) that may be described as a film texture. Figure reproduced from (Morris and Serpell, 2010).

Diffraction from the cross-ß architecture

Figure 3.16 shows a basic fibre diffraction experiment and XRFD pattern from a self-assembled amyloid fibril sample. The pattern exhibits four-quadrant symmetry with reflections occurring on two principle axes, the vertical axis – the meridian pertaining to dimensions on fibre axis and the equator pertaining to dimensions perpendicular to the fibre axis. The interpretation of the cross-ß pattern has been discussed previously (see Section 1.1). Briefly the reflection at ~4.7 Å arises from ß-strands repetitively stacked perpendicular to the fibre axis extending into ß-sheets running along the fibre axis. The ~10 Å reflection arises from the lateral association of ß-sheets perpendicular to the fibre axis.

The factors of disorder and crystallinity affect fibre diffraction patterns in a number of ways. The degree of disorientation affects the degree of radial averaging observed in an XRFD pattern and can be measured as the angular distribution of signals from the major pattern axes. It is essential that disorientation and radial averaging are minimised so that reflections can be assigned to specific unit cell dimensions. The second potential introduction of disorder arises from cylindrical averaging of the fibril sample. As shown in Figure 3.17 this is where the crystallites of a fibril sample may be rotated at random orientations about the fibre axis (Figure 3.17b).

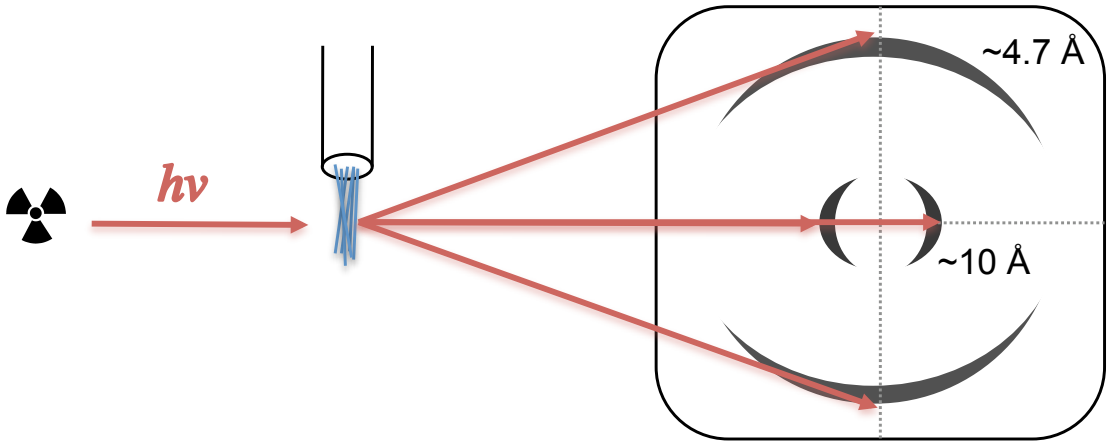


Figure 3.16: A typical fibre diffraction experiment from an alignment of amyloid exhibiting the classical cross-ß pattern.

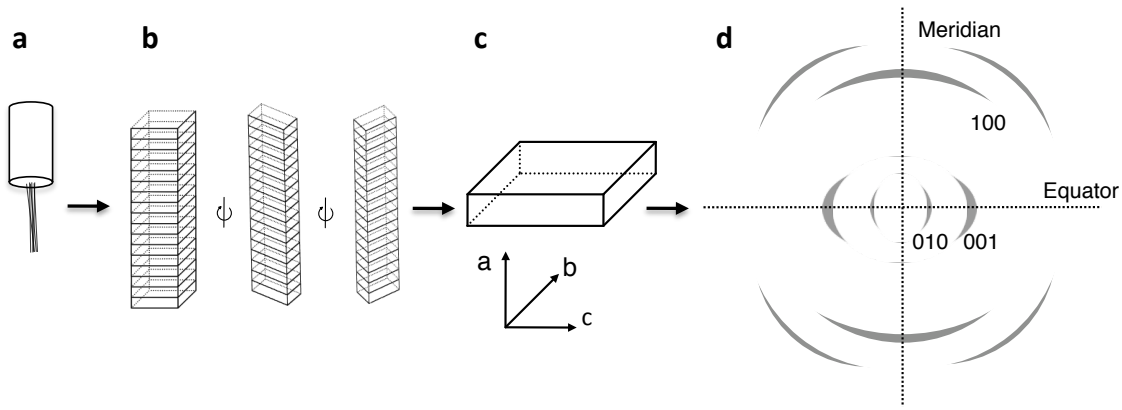


Figure 3.17: General fibre diffraction theory: a bundle of fibres may be aligned by their long axes (a), the fibres may be likened to anisotropic microcrystals which may rotate about their long-axes in the sample (b) the dimensions of the unit cell of the microcrystals (c) are directly relatable to the positions of the reflections on the fibre diffraction pattern (d). Figure reproduced from (Morris and Serpell, 2012).

This results in an inability to unambiguously distinguish between unit cell dimensions and interatomic distances perpendicular to the fibre axis, in other words structural information along the dimensions *b* and *c* (Figure 3.17c). Normally in a fibre diffraction pattern, as illustrated in Figure 3.17d, reflections pertaining to structural features on the *a* dimension are found on the vertical axis of the pattern, that is the axis parallel to the fibre axis or meridian as it is correctly referred to. In a cylindrically averaged sample, reflections found on the equator arise from a mixture of structural information perpendicular to the fibre axis on the *b* and *c* dimensions. This makes it necessary to deconvolute these signals; a number of tools have been developed to do this, described in subsequent sections.

Sample hydration

The use of a humidity chamber in fibre alignments has been reported, this in theory could be used for any type of fibre alignment but has to date been reported to have been successfully used in experiments utilising a fibre textured sample (McDonald et al., 2008). The dehydrated state of structures elucidated from the interpretation of fibre diffraction patterns has been the subject of question recently, a concern being that by dehydrating an amyloid structure one is introducing artefactual structures (Kishimoto et al., 2004, Maurstad et al., 2009). Evidence has

also been presented showing that the structures present in the dehydrated amyloid state are representative of the solution structure (Squires et al., 2006). Regardless of the debate over representative structures, the investigations around this subject have led to the collection of data from semi-hydrated fibre alignments and in these instances more information and apparent 'order' has been observed (McDonald et al., 2008).

Software, analysis and simulation of x-ray fibre diffraction patterns

A number of methodologies exist for analysing fibre diffraction data and a number of analysis packages exist to aid this (Fit2d (Hammersley, 1997), Clearer (Makin et al., 2007), WCEN (Bian et al., 2006)). The choice of analysis method depends largely on the quality of data obtained. A fibre diffraction pattern may only show the classical cross-ß signals. In this case further interpretation is unlikely to be possible but by a comparison of a large number of amyloid systems it is clear that this represents the amyloid architecture and so is useful to obtain (Jahn et al., 2010). If a pattern contains more information i.e. multiple discrete signals on unique axes then it may be possible to interpret more detailed structural models.

The modelling process of XRFD pattern interpretation process can work in a number of ways but to model *de novo* the fibre diffraction pattern must first be indexed, that is assigning diffraction signals to dimensions they arise from and by what order. Manual pattern indexing and interpretation is based on the fact that all of the reflections arising from a unit cell dimension must be related to one another by some multiple. This may be done manually, especially where something is known about the system in question, i.e. chain length, probable intersheet distances, proximity of chromophores, however this process is quite labour intensive. The analysis package Clearer may be used to automatically index diffraction signals and predict likely unit cell dimensions (Makin et al., 2007).

Models may then be built into the unit cells based on the likely number of molecules per cell, knowledge of possible architectures based on other experimental data and stereochemical

considerations. Models may be built in packages such as InsightII (Accelrys) or Chimera (Pettersen et al., 2004). Potential models built into predicted cells should be tested for accuracy by comparison of simulated fibre diffraction to experimental fibre diffraction. Clearer can be used to simulate fibre diffraction patterns as well as the package Disorder (Borovinskiy, 2006). Further modelling and refinement generally then follows an iterative process based on changes to the models and further simulation of X-ray fibre diffraction patterns. Information from other experimental sources may provide useful constraints to produce models that are the best representation of the fibrillar structure being determined.

4 Materials and Methods

4.1 General

All water used was MilliQ filtered water with a resistivity of 18.2 MΩ, 5 ppb at 25 C^o. All balance measurements were made on a microbalance (Denver Instruments SI-234) with a tolerance of ±0.1 mg. All pH measurements were made using a FC200 pH probe (HANNA instruments) with a (6 mm x 10 mm) conical tip to an accuracy of ±0.1. All sterilisation filters (0.2 μm) used were the Millex brand manufactured by Millipore. Where material was used that could cross contaminate cuvettes and surfaces, a cleaning routine of water, 2% Hellmanex, water, ethanol following by drying with $N_{2(}g)$ was used. Where necessary, thorough cleaning of cuvette surfaces was achieved by the use of 2 M nitric acid baths. Wherever possible low-bind plasticware was used to minimise sample loss and binding of fibrillar material. In these cases microfuge tubes were Eppendorf low-bind sterile (Eppendorf, Hamburg, Germany) and Avantguard low-bind sterile filter barrier tips.

4.2 Methods relevant to Chapter 5

4.2.1 Waltz peptide preparation

The Waltz peptides HYFNIF, RVFNIM and VIYKI were synthesised as lyophilised powder with N-acetylated and C-amidated termini using the Fmoc protection scheme. Purity was >95% achieved by HPLC and MS (JPT peptide technologies, Germany). Peptides for CD analysis were synthesised with a trifluoroacetic acid counter ion and for all other analyses with a chloride counter ion. To form fibrils, peptides were incubated in water at 10 $mgmL^{-1}$ under quiescent conditions at room temperature for one week unless otherwise stated. Fibrils allowed to assemble for periods longer than one week are referred to as mature.

4.2.2 Transmission electron microscopy

Material preparation

All instances of grid(s) refer to formvar-carbon film coated 400-mesh copper grids (Agar scientific).

Basic grid preparation

Samples for inspection by transmission electron microscopy (TEM) were prepared on grids by allowing a 4 – 5 µL solution of material to adsorb to the grid surface by a one minute incubation, excess solution was blotted away with grade I filter paper (Whatman). Grids were washed by one-minute incubation with 0.2 µm filtered milliQ water, excess blotted away, and subsequently stained and blotted by two one minute incubations with 2% w/v uranyl acetate (Sigma-Aldrich).

Visualisation

Material prepared for microscopy was inspected using a Hitachi 7100 TEM (Hitachi) operated at 100 kV. Images were acquired digitally using an axially mounted Gatan Ultrascan 1000 CCD camera (Gatan, Oxford, UK).

Image analysis

Raw camera data stored as .dm3 files were converted to .tiff files at 2048 x 2048 pixels. Rapid fibril morphology analysis was performed using macros (see Appendix IV) written in ImageJ (NIH), briefly a user specified line is expanded to select the fibril in question and copied into a new image where a fast Fourier transform (FFT) bandpass filter is used to remove upper and lower frequency data and visualise morphology more effectively.

4.2.3 Circular dichroism

Data collection and plotting

A Jasco J-715 spectropolarimeter (Jasco UK, Great Dunmow, UK) with a peltier temperature control system was used to collect circular dichroism (CD) spectra. All measurements were made at 20 ^{o}C with a continuous scan at a pitch of 0.1 nm at a scan speed of 50 nmmin^{-1} (response time 4 sec, slit width 1 nm). Spectra were standardly measured between 320 – 180 nm. Control buffer spectra were averaged from triplicate scans and subtracted from the averaged triplicate scans of samples. Pathlengths were adjusted according to the degree of scattering vs signal intensity, but generally pathlengths were between 0.1 – 0.01 mm using quartz demountable cells (Starna Scientific Ltd) to ensure good quality spectra down to at least 190 nm. The highest quality spectra for HYFNIF and VIYKI were collected in a 0.01 mm cuvette at 1 mgmL^{-1} whilst RVFNIM was collected at 2 mgmL^{-1} in 0.01 mm. Units are plotted in molar ellipticity with the units deg cm^{2} dmol^{-1}, as calculated using Equation 4.1 where the CD ellipticity signal at a particular wavelength (θ_λ) is scaled proportionally to sample molarity (M) and light pathlength (d).

$$[\theta]_{molar/\lambda} = \frac{100 \cdot \theta_\lambda}{M \cdot d} \qquad \textit{Equation 4.1}$$

Whilst path length and concentration was adjusted according to the degree of scattering vs. signal intensity to ensure good quality spectra, CD data recorded with a high tension voltage (HT[V]) of >600 was automatically removed by data-processing in Excel. This ensures no artefacts in the spectra are reported.

Linear dichroism artefact identification

To check for contributions arising from linear dichroism (LD) artefacts due to sample orientation effects, cuvettes were mounted close to the detector and scans taken at 0 and 90°. By mounting the cuvette close to the detector this also checks for chiral scattering effects. Where

artefacts arising from orientation of fibrils were found, a combination of tip (Sonics Vibra-Cell VCX500, 20 kHz, 40% amplitude, 2 mins) and water bath (Fisher Scientific FB15051, 37 kHz, 1 min) sonication were used to disrupt alignment (Andersen et al., 2010).

Data analysis and dichroweb

Secondary structure analysis was performed on spectra confirmed to contain only CD signals using the online server Dichroweb (Whitmore and Wallace, 2004) with the CDSSTR analysis programme (Compton and Johnson, 1986) and the SP175 190 – 240 nm reference set (Lees et al., 2006). Careful attention was paid to the closeness of fit between secondary structure model predictions and the experimental data where low NRMSD (<0.1) values indicate a high goodness of fit of the secondary structure model prediction and indication of successful analysis (Mao et al., 1982, Whitmore and Wallace, 2004). Further analysis by Gaussian deconvolution of spectra was performed in Microsoft Excel using Solver. Accurate peak intensity and positions were found by modelling signals using the Gaussian distribution function shown in Equation 4.2, where Gaussian peak intensity (y) is proportional to peak intensity (a_1), wavelength (x), peak position (c_1) and peak width (w_1), and minimising the sum of the square residuals between the total model and experimental data.

$$y = a_1 e^{-\frac{(x-c_1)^2}{(2w_1)^2}} \qquad \textit{Equation 4.2}$$

4.2.4 Linear dichroism

A Jasco J-815 spectropolarimeter (Jasco UK) modified for linear dichroism was used to collect LD spectra. Measurements were taken at room temperature (20 °C) between 320 – 180 nm with a pitch of 0.2 nm at a scan speed of 100 nmmin^{-1}, a response time of 1 second, slit width of 1 nm and standard sensitivity on samples of a concentration of 200 µgmL^{-1} to resolve far and near-UV peaks in a single wavelength scan. Spectra were collected as an average of three accumulated measurements and buffer subtracted. Three channels were monitored: LD, high

tension voltage (HT[V]) and isotropic absorbance. The isotropic absorbance channel is recorded by the conversion of the HT[V] signal to Abs based on empirical calibration of the relationship between HT[V] and Abs on the individual instrument. To unambiguously detect LD signals fibrillar samples were aligned by a couette flow system built in house (equivalent models available from Kromatek, Great Dunmow, UK) (Dafforn et al., 2004) using rotation speeds of 1500 – 3000 rpm, assembly of the couette apparatus was found to align fibres perpendicular to the flow direction (orientation axis) and so comparison spectra were also collected with no rotation. Where spectra show a range of couette rotation speeds care was taken to subject samples to the lowest rotation speed and then subsequent higher speeds. Data are reported in differential absorbance units (ΔAbs).

Chromophore orientation determination

To minimise errors due to signal artefacts from excessive light scattering or absorbance, data where the HT[V] exceeds 600 V is excluded from analysis. To correct the LD spectra prior to chromophore orientation determination, briefly the data was corrected for scattering and the 240 – 190 nm deconvoluted. Scattering correction was performed by fitting a scattering profile, to the non-absorbing region of 320 – 290 nm, with the model shown in Equation 4.3 where background scattering (LD^{τ}) at a particular wavelength is proportional to a scaling factor (a), wavelength (λ) and the model exponent ($-k$), as described previously (Nordh et al., 1986). The model was fit using Solver within Excel and then extended over the full wavelength range after which it could be subtracted from LD spectrum.

$$LD^{\tau}(\lambda) = a.\lambda^{-k} \qquad \textit{Equation 4.3}$$

To account for signal position and intensity shifting due to overlapping peaks the 240-190 nm region was deconvoluted using a three-peak Gaussian distribution model and minimisation by least squares residuals as detailed in Section 4.2.3 and Equation 4.2. The orientation of chromophores was then calculated by calculating the degree of alignment, S, in Equation 3.3

from known chromophore orientations. Using the calculated value for *S,* the angles of individual transitions were calculated from measured LD peak intensities using Equation 4.4 where the angle between the orientation axis and the transition polarisation (α) is proportional to the reduced LD signal (LD^r) and the degree of orientation (S).

$$\alpha = \cos^{-1}\sqrt{\left(\frac{LD^r}{1.5S}+1\right)\Big/3} \qquad \textit{Equation 4.4}$$

4.2.5 X-ray fibre diffraction & analysis

Fibril alignment

Fibrils were aligned by a number of methods, see Chapter 3, to create different textures:

Fibre texture:

A 10 µL droplet of 10 mgmL^{-1} fibril solution was suspended between two wax-tipped 1.2 mm O.D, 0.94 mm I.D borosilicate capillaries (Harvard apparatus) and placed in a parafilm sealed petri dish at room temperature. Care was taken to select wax-tipped 1.2 mm O.D, 0.94 mm I.D borosilicate capillaries (Harvard apparatus) that had a very slight concave surface but not overly so or convex. The suspension was allowed to dry by evaporation. Difficult to align samples were sealed or incubated at 4 °C to slow drying and aid alignment.

Film texture:

A 10 µL droplet of 10 mgmL^{-1} fibril solution was dried onto a clean Teflon surface. The mat was subsequently removed with a scalpel and carefully mounted with superglue onto thin capillaries (Harvard apparatus) created by drawing through a Bunsen flame. Alternatively a 10-50 µL droplet of 10 mgmL^{-1} fibril solution was drawn into a X-ray transmissible 0.7 mm borosilicate capillary (Capillary Tube Suppliers Ltd) and sealed at one end to prevent further capillary action. This solution was allowed to dry by evaporation to a disc texture.

Fibres were routinely inspected by cross-polarisation microscopy prior to X-ray fibre diffraction for record keeping and indication of the likely success of X-ray fibre diffraction.

Data collection

Standard

For fibrous textured alignments, samples were mounted and aligned on the goniometer head ensuring the beam axis was perpendicular to the fibre axis. Exposures were taken at 0 and 90° (Φ) rotation (0.5° oscillation) about the fibre axis to ensure cylindrical averaging. For film textured alignments, samples were mounted and aligned on the goniometer head ensuring the beam axis was parallel to the plane of the film. Exposures could be taken at 0 and 90° (Φ) rotation (0.5° oscillation) achieving beam shots parallel and perpendicular to the film plane.

Real-time alignment

Real-time alignment X-ray fibre diffraction analysis was achieved by the use of a humidity control chamber (as similarly described in (McDonald et al., 2008)). Samples of 10 $mgmL^{-1}$ fibril solutions were suspended between wax-tipped capillaries in the alignment chamber as shown in Figure 4.1. Alignment was achieved in the absence of direct line of sight to the sample by alignment to the 'alignment pin' and subsequent visual adjustment of the goniometer height to bring the sample into the beam. X-ray exposures were started immediately after sample drop suspension at an interval of 10 minutes.

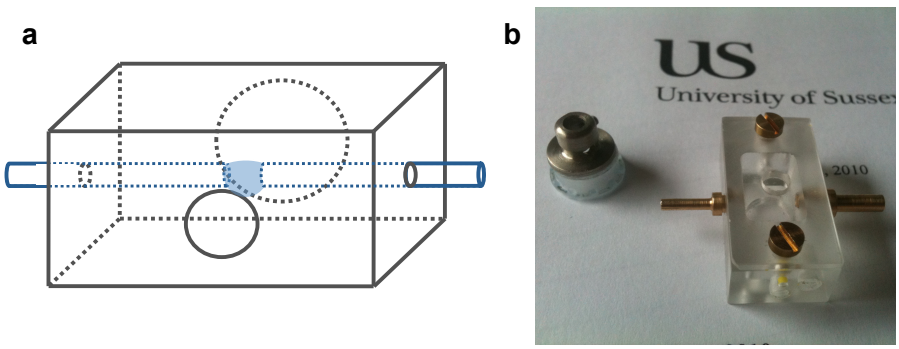


Figure 4.1: The humidity control chamber used for real-time alignment X-ray fibre diffraction analysis, schematic (a) and the chamber showing mountings and 'alignment pin' (b).

X-ray sources and detectors

Rigaku – RAxis (home source)

Data was collected using a Rigaku Cu Kα (λ 1.5419 Å) rotating anode generator and RAxis IV++ plate detector with exposure times of 10 – 60 minutes and specimen to detector distances of 160 or 250 mm. The detector recorded images at 3000 x 3000 px at 100 μmpx^{-1}, whilst images for processing were compressed to 752 x 752 px at 400 μmpx^{-1}.

Rigaku - Saturn 944+ (home source)

Data was collected using a Rigaku 007HF Cu Kα (λ 1.5419 Å) rotating anode generator with VariMax-HF mirrors and Saturn 944+ CCD detector. Exposure times were typically 10 – 120 seconds and specimen to detector distances 50 or 100 mm. The detector recorded images at 1042 x 1042 px in (2 *x* 2 binning mode) at 90 μmpx^{-1}, whilst images for processing were compressed to 523 x 523 px at 179.3 μmpx^{-1}.

Diamond – MarCCD (synchrotron source)

Data was collected at the I24 MX microfocus beamline at the Diamond synchrotron radiation facility (Oxfordshire, UK) at a wavelength of 0.9778 Å and a MarCCD CCD detector. Exposure times were typically at 50% attenuated intensity for 1 – 5 seconds and specimen to detector distances were 312.2 mm unless otherwise stated. The detector recorded images at 4099 x 4099 px at 73.2 μmpx^{-1}, whilst images for processing were compressed to 684 x 684 px at 439.2 μmpx^{-1}.

Data processing

Diffraction images were opened using iMosflm for initial inspection, for conversion to .tiff files, the mosflm 7.0.5 GUI was used (Leslie, 1992). Data processing was carried out using Clearer (Makin et al., 2007). References to modules within Clearer are made using italicisation. The deviation of the meridional angle from vertical was measured using the *Radially Average* module and the pattern rotated so that this was vertical, following which patterns were cropped

their original size by the *Crop to Central Square* command. Patterns were centred using the *Centre* module, which compares the intensity profile from the meridional and equatorial pattern axes to align them, assuming four-quadrant symmetry in fibre diffraction patterns. Centred patterns were also checked using the *Radially Average* module as shown in Figure 4.2.

Prior to signal position measurements, the parameters described earlier (X-ray Sources and Detectors) were entered in the *Diffraction Settings* module. Automated signal position measurements were made using *Peak Find* within the *Radially Average* module with a 60° angular search width. The sampling interval (*Peak Search Width*) was altered to adjust search sensitivity as required per pattern. Clearer outputs the systematic errors associated with measured signals as a function of pixels, these values were converted to Å by Bragg's law implemented in Excel. This method does not compensate for flat detectors and so is used only for graphical representation and comparison of patterns collected using the same detector and diffraction settings. All signal positions were confirmed using the *Zoom and Measure* module and any signals measured entirely manually are indicated so by the use of *.

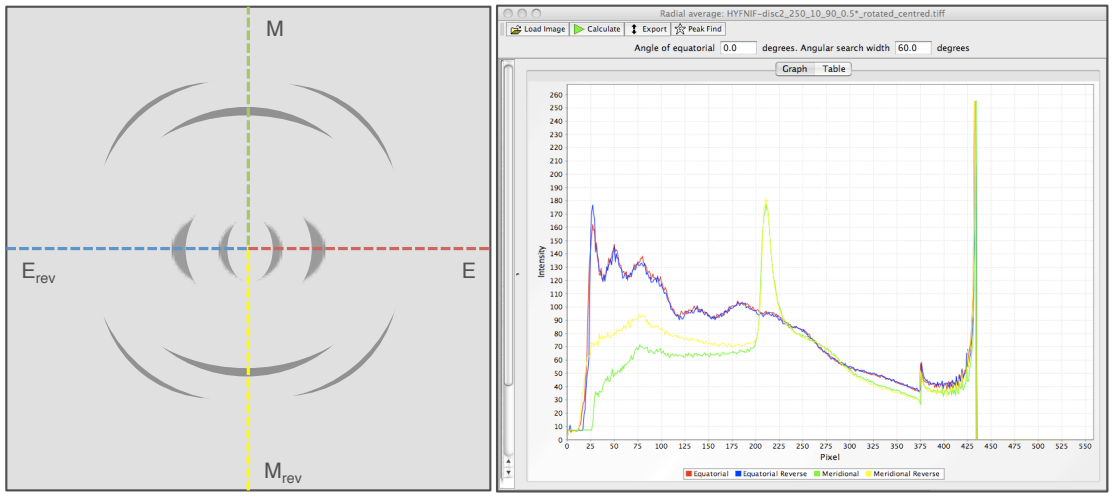


Figure 4.2: X-ray fibre diffraction patterns are divided into four symmetrical quadrants separated by the meridional and equatorial axes (a), the intensity profile along the equatorial (E) and meridional (M) may respectively be compared to the equatorial reverse (E_{rev}) and meridional reverse (M_{rev}) to centre a fibre diffraction pattern (b) in the program Clearer (Makin et al., 2007).

Unit cell determination

The most likely orthorhombic unit cell dimensions were explored and determined using the *Unit Cell Optimisation* module. Initial unit cell guesses (herein referred to unit cell search limit) were made based on the largest and strongest signal positions as well as the peptide size. Where cross-ß patterns were observed, the fibre axis dimension was not explored and assumed to be equivalent to the major meridional reflection of the pattern being indexed. The perpendicular dimensions were explored in the unit cell determination process. Signal positions were entered with minimum and maximum Millar indices calculated according to the unit cell search limits as specified in the relevant sections.

Potential unit cells identified by the *Unit Cell Optimisation* module are rated by comparison of the back calculated signal positions to the experimental signal positions producing an error value. Low error unit cells identified were preferentially explored in the modelling process.

4.2.6 Modelling and X-ray fibre diffraction simulation

Modelling from X-ray fibre diffraction was conducted in an iterative manner but, considered as linear steps, is shown in Figure 4.3 and described below:

Modelling

Initial peptide models were built in Insight II (Accelrys) in an ideal ß-strand conformation. Initial rotamer conformations were assigned using a backbone dependent rotamer library (Dunbrack and Cohen, 1997). The dimensions and volume of the modelled peptide were determined using the PDB co-ordinates, Excel and Chimera (Pettersen et al., 2004), and comparison to the predicted unit cell was used to estimate the number of peptides per unit cell. Preliminary peptide arrangements were explored by construction in Pymol (DeLano, 2002) based on visual consideration of steric overlaps. Molprobity (Davis et al., 2004, Lovell et al., 2003) was used to assess constructed models on the basis of steric overlaps, $\Phi - \Psi$ angles and

rotamer conformation. Models were minimised using NAMD within VMD (Phillips et al., 2005) and the CHARMM Prot_all_22 forcefield (Brooks et al., 2009). Briefly, fibres were constructed from initial models using Pymol scripts (see Appendix IV) and determined unit cells. These were solvated in a water box extending 10 Å beyond each model dimension and minimisation subsequently run over approximately 500 cycles. Peptide models were extracted from the centre of the minimised fibre models and analysed by fibre diffraction simulation.

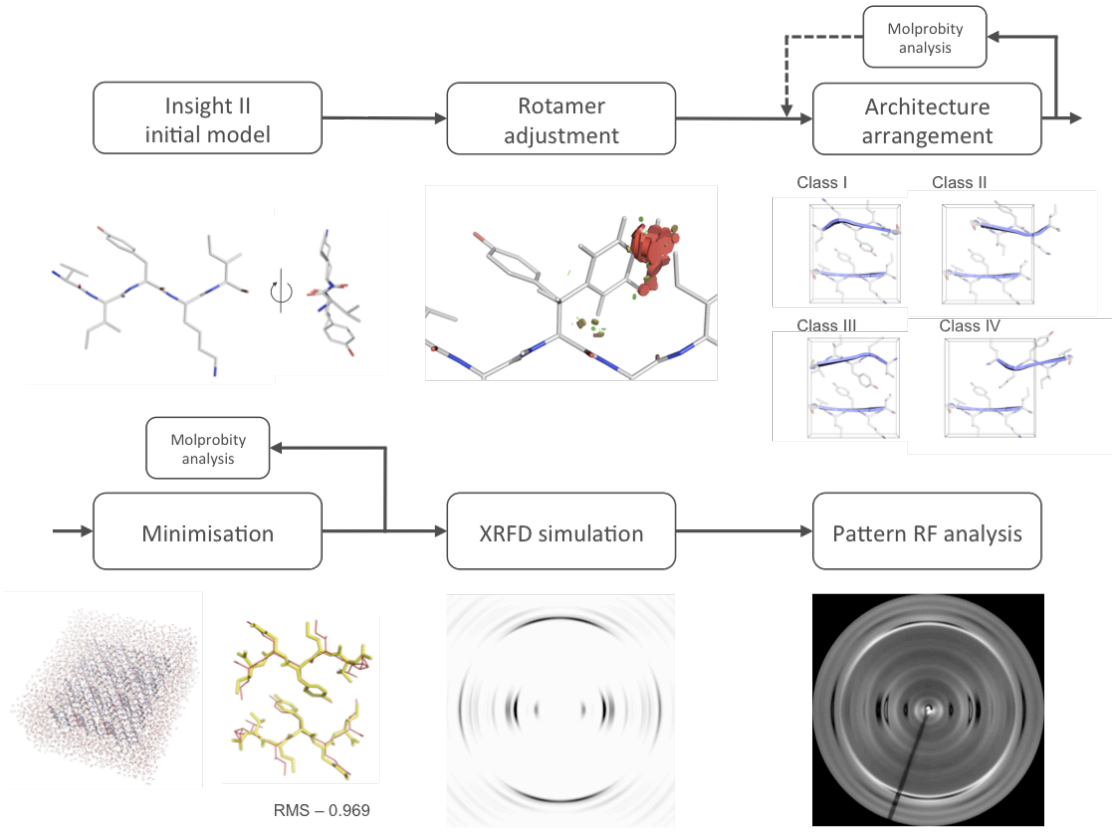


Figure 4.3: Model building schematic showing general workflow from unit cell prediction.

Fibre diffraction simulation

Models were inspected in the fibrous texture using the *Structure Chain Generator* module and unit cell dimensions. Fibre diffraction patterns were simulated from PDB models using the *Fibre Diffraction Simulation* module of Clearer. Simulation settings were input to match the experimental pattern being compared to (see Section 4.2.5). For orthorhombic unit cells the fibre axis Cartesian co-ordinates were set to [001] corresponding to the *c* dimension used by convention here for the fibre axis direction. Fibre disorder was 0.2 radians with a sampling interval of 1 pixel. Contrast was changed according to the best visualisation of the pattern signals. Crystallite size was increased to between 400 – 800 Å. Detector size was adjusted to the size of the experimental .tiff pattern. All other parameters were set to default values.

4.2.7 Fibre diffraction pattern comparison and analysis

Several methods of comparing simulated and experimental fibre diffraction patterns were employed to identify the model structures that best represented the Waltz fibrillar architectures:

Qualitative - visual

Assuming the pixel size and detector size of the simulated and experimental fibre diffraction patterns were matched, it is possible to overlay two patterns and reduce the size the simulated pattern to one of the four quadrants of the experimental pattern for visual comparison of the equatorial and meridional signal. The *Layers* module of Clearer may also be used to overlay two patterns, which are pseudo coloured such that the overlay outputs a type of interference map to identify signal matches. This allows a good initial assessment of the 'goodness of fit'.

Qualitative - graphical

Further qualitative assessment was made by plotting the equatorial and meridional signal positions, widths and intensities as a graphical trace within Excel using Bragg's law (Equation

3.5). Signal intensity information as a function of distance in pixels was extracted from the *Radially Average* module of Clearer.

Quantitative

MATLAB (The Mathworks, Inc) was used to perform full quantitative comparative analyses of simulated and experimental diffraction patterns of the same dimensions and pixel size. Prior to analysis, patterns were cropped to a central circle to include only resolutions greater than 3 Å. This was done so to increase the pixel signal to noise ratio and improve comparison of experimental and simulated patterns. The MATLAB script for pattern comparison can be found in the Appendix IV. In summary, the diffraction images to be compared were stored as pixel intensity matrices.

The intensity of each pixel of the two arrays was converted to a decimal value and compared individually. The difference values were stored in a new matrix to create a difference map as shown visually in Figure 4.4. Low pixel difference values (black) indicate high correlation where high pixel difference values (white) indicate low correlation. The absolute difference between the pixel matrices is calculated by summation. The RF value is expressed as an average of this total difference.

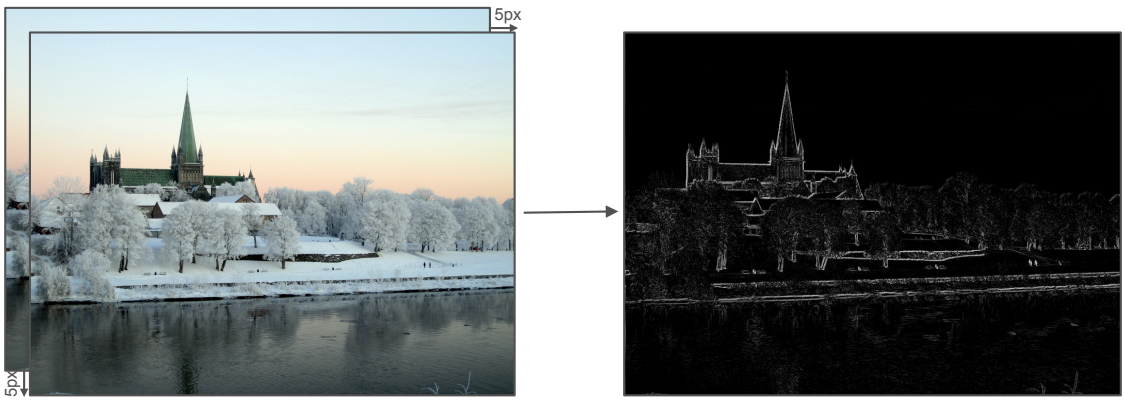


Figure 4.4: Demonstration of the MATLAB RF analysis script. Two images shifted by a minimal number of pixels may be compared to one another resulting in a subtle difference map where black denotes high correlation and white low correlation.

4.3 Methods relevant to Chapter 6

4.3.1 Sample preparation

Unmodified N-VLYVGSKT-C peptide (Advanced Biomedical Ltd, Oldham, UK) was dissolved at 5 $mgmL^{-1}$ in water. The concentration of the dissolved solution was increased by rotary evaporation at 40 – 45 °C for a total of 90 minutes with regular inspection of the sample volume. The peptide concentration was calculated to be 150 – 200 $mgmL^{-1}$ based upon solvent reduction.

4.3.2 Transmission electron microscopy

Materials used, basic grid preparation, visualisation and image analysis were performed as in Section 4.2.2.

4.3.3 X-ray fibre diffraction & analysis

X-ray fibre diffraction, processing and analyses were performed as described in Section 4.2.5.

4.3.4 Modelling and X-ray fibre diffraction simulation

Models of α-synuclein ß1 were built using Chimera in an ideal parallel ß-sheet geometry and subsequently minimised within the same package (Pettersen et al., 2004) as briefly described. Using the unit cell dimensions a [3 3 3] lattice was created such that the innermost unit cell was interfaced with a neighbouring unit cell on each dimension. Hydrogens were added and charges assigned using the integrated MMTK (Hinsen, 2000) and the AMBER ff99SB forcefield (Wang et al., 2000). The steepest decent minimisation method used was used over 100 steps at a step size of 0.02 Å. X-ray fibre diffraction simulation was subsequently performed on constructed models as described in Section 4.2.6.

4.4 Methods relevant to Chapter 7 – 8

4.4.1 Naphthalene dipeptide preparation

Dipeptide dissolution

Naphthalene conjugated dipeptides were synthesised by Dr. D. J. Adams (University of Liverpool) as described in detail previously (Chen et al., 2010a, Chen et al., 2010c). Briefly, naphthalene dipeptides were prepared from 2-napthol, 6-bromo-2-napthol or 6-cyano-2-napthol and L-amino acids added sequentially to yield the final naphthalene conjugated dipeptide product. Molecular identity and purity were assessed by ^{1}H and ^{13}C NMR using a Bruker Avance 400 NMR spectrometer operated at 400.12 and 100.60 MHz respectively and mass spectrometry. Lyophilised dipeptide was dissolved in water at 5 $mgmL^{-1}$ with an equimolar quantity of NaOH to adjust the pH of the solutions to above the pKa of the carboxy-terminus, between 10.7 – 10.5 as stated per chapter. Solutions were incubated at 37 °C and agitated at between 1000 – 1500 rpm overnight to fully dissolve the dipeptide and allow pH equilibration. Fine adjustment of pH was achieved using 100 mM HCl or NaOH. Care was taken to not adjust the volume of the solution by more than 1%, maintaining accuracy of the dipeptide concentration.

Dipeptide gelation

Solutions of dipeptide were mixed with GdL at specific concentrations to achieve the desired final pH, as stated per chapter. The GdL was preweighed and the dipeptide solution added to this. Care was taken to ensure gentle but thorough mixing. For time course measurements the time between GdL dissolution and initial data collection was less than 2 minutes.

4.4.2 General considerations

The principle difference in the structural and biophysical characterisation of dipeptide hydrogels relates to the method of sampling. It was established that sampling from pre-formed gels by pipetting was not effective for removing the gelating material where it was likely sequestered

within gel matrices (see Chapter 7). As such to assay the gel material dipeptides for analysis were prepared *in situ*.

4.4.3 Transmission electron microscopy

Materials used, basic grid preparation, visualisation and image analysis were performed as in Section 4.2.2. Methods specific to TEM of the dipeptides is detailed as follows:

In situ grid preparation

In situ – time course humid chamber

Dipeptide assembly was initiated by GdL pH reduction as described. These solutions were immediately transferred onto parafilm (Bemis) in a humid chamber at room temperature. Grids were inverted onto 10 μL drops of gelating solution, incubated for set periods of time and removed when appropriate dried and stained twice by one minute incubations with 2% w/v uranyl acetate (Sigma-Aldrich).

In situ – dry down

For visualisation of final fibril structure, 1 μL aliquots of dipeptide were assembled *in situ* on upright grids in a humid chamber at room temperature overnight. The gelating solution was allowed to dry completely and subsequently stained twice by one minute incubations with 2% w/v uranyl acetate (Sigma-Aldrich). At the edges of the dried down gel drop, material could be inspected. This method was also employed for the analysis of high pH stocks.

Micrograph analysis

TEM micrographs were analysed using ImageJ (Abramoff et al., 2004). Minimum diameter fibril were identified by careful visual comparison. Fibril diameters were measured by making at least two measurements per fibril identified. Intra and interfibril variation is reported as a single standard deviation of measurement, with the number of measurements made as ‘n’.

4.4.4 UV/Vis

UV/Vis spectra were recorded using a Shimadzu UV-2401 PC spectrophotometer. Absorbance data was collected between 800 – 190 nm with a resolution of 1 nm at room temperature using a 10 mm pathlength quartz cuvette (Hellma #109.004-QS). Spectra were recorded for stocks of naphthalene dipeptides prepared at 5 $mgmL^{-1}$ at a pH of 10.5. Prior to data collection these were diluted to 25 and 3.13 μgmL^{-1} for near and far-UV respectively.

4.4.5 Circular dichroism

CD spectra collection and LD artefact identification was performed as in Section 4.2.3 but at a temperature of 25 (Chapter 7) and 21 ^{o}C (Chapter 8) between 350 – 180 nm. Control buffer spectra were water where the contribution from NaOH is negligible. Subtraction of the spectra from GdL was not included in data processing where it was found to be variable over time but its contribution negligible. Where extended time course measurements were performed, demountable cuvettes were parafilm sealed to minimise evaporation. The interval for measurements was 10 minutes unless otherwise stated. To calculate the absorbance profile from assembling dipeptides the $HT[V]_{sub}$ was calculated. This is derived from the HT[V] of triplicate scans of buffer subtracted from triplicate scans of the HT[V] of the sample.

4.4.6 Fluorometry

Fluorescence measurements were made using a Cary Eclipse spectrofluorometer (Varian) with a single cell Peltier accessory set to 25 °C (Chapter 7). The naphthyl fluorescence was measured using an emission scan at 200 nm min^{-1} with slit widths at 5 nm over 500 – 280 nm with an excitation wavelength of 265 nm.

4.4.7 X-ray fibre diffraction & analysis

X-ray fibre diffraction pattern collection and analysis was performed as in Section 4.2.5. Methods specific to XRFD of dipeptides are as follows:

In situ fibril alignment

Dipeptide fibre alignments were prepared as in Section 4.2.5 but using a freshly gelating solution prepared as described in Section 4.4.1.

Dehydration studies

Aligned dipeptide fibres were subjected to dehydration by vacuum desiccation for extended periods of time. Dehydrated fibrils were rehydrated by inverted suspension in a water-filled Eppendorf for one-minute intervals. During the data processing and analysis of variably hydrated fibre diffraction patterns contrast ratios were maintained between patterns so that the relative intensities of reflections between patterns could be compared.

4.4.8 Modelling and X-ray fibre diffraction simulation

X-ray fibre diffraction simulation was subsequently performed on structures as described in Section 4.2.6. Where the crystal structure of 2-AV is in a monoclinic cell, additional procedures were necessary to specify the correct coordinates of the fibre axis. The following describe this procedure assuming *c* is the fibre axis and ß ≠ 90: The monoclinic cell dimensions were specified including the non-90° ß angle. The *Real Space* Cartesian co-ordinates *ljk* were noted for the *c* dimension corresponding to the fibre axis. The *Beam Orientation, Beam Vector* was set to either [100] or [010] perpendicular to the fibre axis and subsequently the *Real Space* Cartesian co-ordinates, *ljk*, entered as the *Fibre Axis* parameters.

4.4.9 Dipeptide mixed gelation

Br-AV and 2-AA were independently aligned *in situ* by starting gelation with the addition of GdL (4 $mgmL^{-1}$) to a stock of the di-peptide (5 $mgmL^{-1}$). A 10 µL drop of this solution was suspended between two wax tipped capillaries and sealed in a petri dish. The solution was then allowed to dry and the fibres align. A mixture of Br-AV and 2-AA were aligned in the same way at a total dipeptide concentration of 5 $mgmL^{-1}$ in a 1:1 ratio using GdL at 8.9 $mgmL^{-1}$.

5 Results and Discussion: Elucidation of the amyloid-like structure of the Waltz short peptides

5.1 Introduction

5.1.1 Amyloidogenic sequences

Just as sequence determines the three-dimensional structure of a natively folded protein in its natural environment, it has long been recognised that sequence affects a protein's propensity to aggregate and self-assemble. Although the ß-sheet rich assembly appears to be a conformation inherently accessible to the peptide backbone (Dobson, 2001), not all un-structured and unfolded proteins aggregate. In the unstructured state, where there is high backbone and residue exposure, a protein sequence is most prone to aggregate and self-assemble (Chiti and Dobson, 2006). Thus the fact that not all proteins do, further suggests some sequences are more amyloidogenic than others (Lopez de la Paz and Serrano, 2004).

Seminal investigations systematically explored the amyloidogenic properties of hexapeptides produced from scanning amino acid mutation of the amyloid-forming fragment STVIIE (Lopez de la Paz and Serrano, 2004). It was established that in ß-sheet forming hexapeptide sequences end positions can accommodate a broader range of amino acid substitutions (including charged residues) but, the core of the peptide sequence is much more restrictive to substitution. This lead to a library of amyloid-forming peptides being created identifying sequence determination rules for self-assembly and the potential to identify dangerous amyloid forming sequences throughout the proteome (Lopez de la Paz and Serrano, 2004). The authors of this study noted that their library represents a small sequence space for amyloid formation and further characterisations of

unrelated sequences would be beneficial. Indeed a number of algorithms and methods, described below, for identifying amyloidogenic sequences have since emerged in a hope to understand the relationship between sequence and amyloid formation propensity and in some cases the rate of fibril formation.

5.1.2 Aggregation propensity prediction

Based on physiochemical analyses of known natively unfolded amyloidogenic proteins it is possible to scan a sequence and find the aggregation-prone regions. These include hydrophobic residues and particular patterns of these along with the propensity to form ß-sheets. This method has been used to reliably predict the increased aggregation rates of pathogenic mutations of amyloid-ß (Aß), including the Artic (E22G), Dutch (E22Q), Italian (E22K) and Iowa (D23N) mutants (Pawar et al., 2005). The concept of aggregation hot-spots was introduced as it was found that short sequences could act to facilitate or inhibit aggregation of polypeptides, presumably through interaction with these hot-spots. Further, these hot-spots have a low incidence in unfolded polypeptides but are readily present folded polypeptides where they can be protected from making intermolecular contacts by the natively folded state (de Groot et al., 2005). Further investigations introduced even more complex sequence position scanning algorithms based on amino acid physiochemistry, aromaticity, ß-sheet propensity, charge, solubility, condition temperature and concentration (Tartaglia et al., 2005). The SALSA (Simple Algorithm for Sliding Averages) program was developed to find aggregation hot-spots in unfolded polypeptide sequences based on ß-sheet propensities, although the authors note the program could be calibrated with other amino acid properties (Zibaee et al., 2007). Conceptually similar programs such as Tango have been developed but which include additional physiochemical properties (Fernandez-Escamilla et al., 2004). As well as ß-sheet propensity alone, it has been found that a sequence's ability to switch between secondary structures influences a polypeptide's ability to aggregate (Hamodrakas et al., 2007).

The methods outlined above make use of physiochemical properties and statistical analysis but another approach adopted for aggregation prediction is through structural modelling. One such method attempts to thread potential sequences onto a range of crystal structures of amyloid-like backbones. Those sequences found to be energetically favourable are considered to have a propensity to form amyloid (Thompson et al., 2006). This method would identify sequences that can self-complement one another in the same manner as the crystal template used for the prediction. The ability to predict amyloid forming segments with consideration of physiochemical properties through the use of the ROSETTADESIGN forcefield highlights the additional energetic and steric influences involved amyloid formation (Thompson et al., 2006). However, this method is presumably limited to successfully predicting only sequences that adopt similar molecular packing to the structural template used. Residue packing density (Galzitskaya et al., 2006) and interfibril hydrophobic interfaces (Saiki et al., 2006) have also been considered in a structural approach to the prediction of amyloid forming sequences.

5.1.3 Ordered aggregation prediction

The methods described are able to predict a subset of amyloid forming segments, but are dependent on sequence-based predictions that discount structural influences or those that do, for the most part, are limited by the small number of experimentally validated amyloid templates. Additionally only the structural-based methods will exclusively predict ordered aggregation (Maurer-Stroh et al., 2010). Physiochemical and energetic considerations have been combined with residue number in a position-scoring matrix to predict amyloid-forming segments based on residue position as well as incorporating the prediction parameters made by other prediction algorithms to make the Waltz algorithm (Maurer-Stroh et al., 2010), as summarised in Figure 5.1.

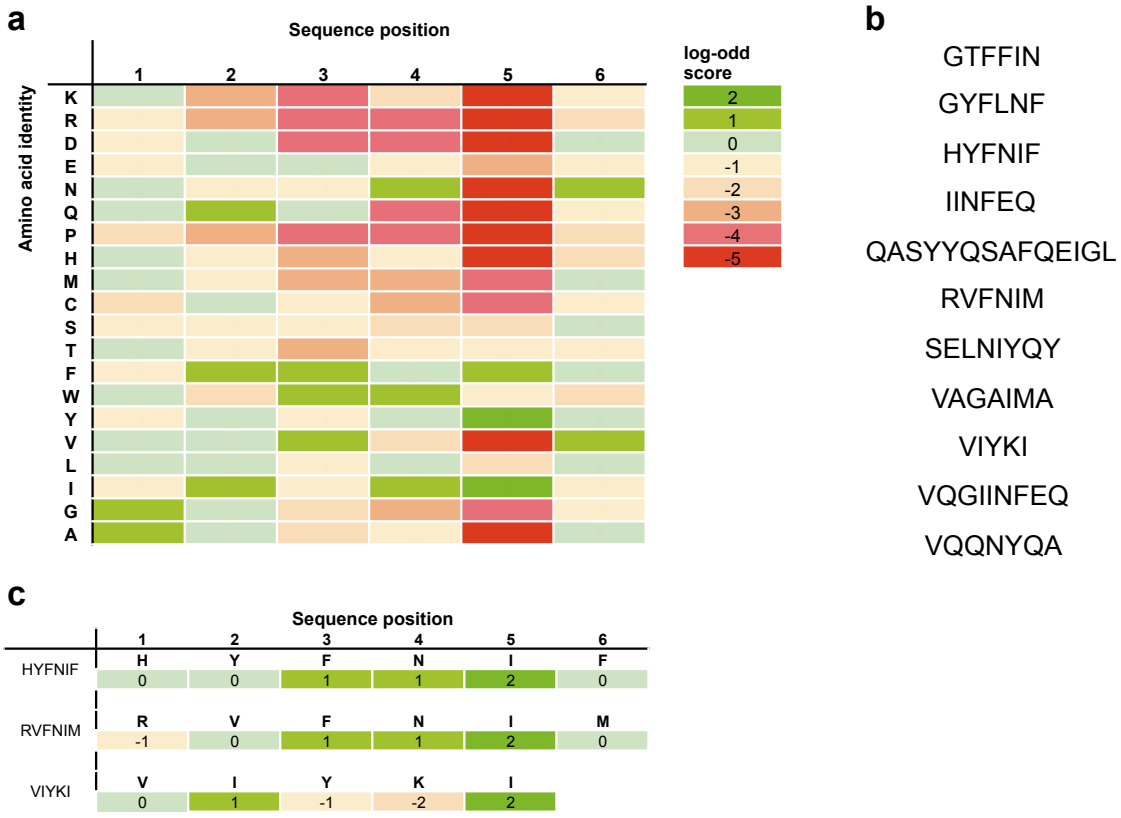


Figure 5.1: The position-specific scoring matrix used by the Waltz algorithm; residues incompatible with a position are denoted by red shading and those compatible with green shading. The log-odd score is a measure of the probability of a residue occurring at a specific position (a). A range of sequences identified in the creation of Waltz are shown (b). The scoring matrix for the three peptides HYFNIF, RVFNIM and VIYKI are also shown (c). Figure adapted from (Maurer-Stroh et al., 2010).

This algorithm aims to distinguish between amorphous and ordered aggregation, but further, to explore the amyloid sequence space to expand the available structures representing this architecture (Maurer-Stroh et al., 2010). Indeed, the algorithm has successfully expanded the sequence diversity of the AmylHex database (Thompson et al., 2006) by almost 400%.

The Waltz algorithm was developed iteratively, and so peptide sequences are available from the training process as well as the predictions (Maurer-Stroh et al., 2010). Twelve of these short peptides were characterised, though presented here are the best structural and biophysical data collected from two hexapeptides (HYFNIF – UniProt ID: P54132, RVFNIM – UniProt ID: P20042) and one (VIYKI – UniProt ID: P07182.2) pentapeptide. All references to the Waltz peptides describe them in their fibrillar form unless otherwise stated and all spatial and

geometrical statements are made with reference relative to the fibre axis. Methods have been reported in the literature that monomerise amyloidogenic proteins and peptides (Broersen et al., 2011), where the results presented here concern the final fibrillar structure of the Waltz peptides these methods have not been employed.

5.1.4 Experimental Aims

The investigations presented here aim to validate the Waltz algorithms ability to predict ordered amyloid forming peptide sequences and where the sequences are novel, new structures may be proposed expanding the representation of the amyloid structural space. The characterisation of new amyloid forming sequences also presents an opportunity to corroborate existing and identify any new basis for amyloid formation. Methodologically, amyloid fibrils present a challenge to structurally characterise and so the methods and considerations of these presented here have broad implications for the investigations of these types of systems.

Further, the in depth characterisation of the amyloid-like peptides here will no doubt serve as a good basis for continued model studies of amyloid formation and disease processes but also self-assembly in general.

This chapter expands on initial results published by Maurer-Stroh *et al.* on the development of the Waltz algorithm (Maurer-Stroh et al., 2010), presenting the analysis of data collected by K. L. Morris on three of the peptides identified. Prof. J. W. H. Schymkowitz and Prof. F. Rousseau (University of Leuven) are gratefully acknowledged for supplying a significant amount of the peptide material. M. Debulpaep (University of Leuven) is gratefully acknowledged for the collection and supply of the Fourier transform infrared spectroscopy data.

5.2 Results and Discussion

5.2.1 Waltz sequences and native protein structure

The Waltz peptides were identified from the full-length native proteins summarised in Table 5.1. Secondary structure prediction was performed on the Waltz peptides in their native full-length sequences using the online server PSIPRED (Buchan et al., 2010, Jones, 1999). RVFNIM and VIYKI were predicted within their native sequences to have helical and ß-strand secondary structure respectively but unfortunately no experimentally determined three-dimensional structures of the full-length native proteins are available for comparison to these predictions. Within its native sequence HYFNIF is predicted to have a ß-strand conformation, however, the sequence adopts an α-helical conformation in its native structure as shown in Figure 5.2d. It is interesting that this prediction is made based on secondary structure propensity of amino acids but a different conformation is adopted natively. This suggests that the three-dimensional context of the sequence safeguards it from its ß-strand propensity, perhaps as a safe guard against aggregation. This concept has been reported in the literature to be evident in the primary sequence of proteins in the form of 'gate keeper' residues (Reumers et al., 2009) but may also be achieved through conformational restriction at levels higher than that of the primary amino acid sequence.

Waltz sequence	UnitProtID	Name	Function	Organism
HYFNIF	P54132	Bloom syndrome protein	DNA helicase	Human
RVFNIM	P20042	eIF-2	Translation regulation	Human
VIYKI	P07182	Chorion protein	Egg shell protein	Drosophila

Table 5.1: The Waltz peptides and their native full-length sequences.

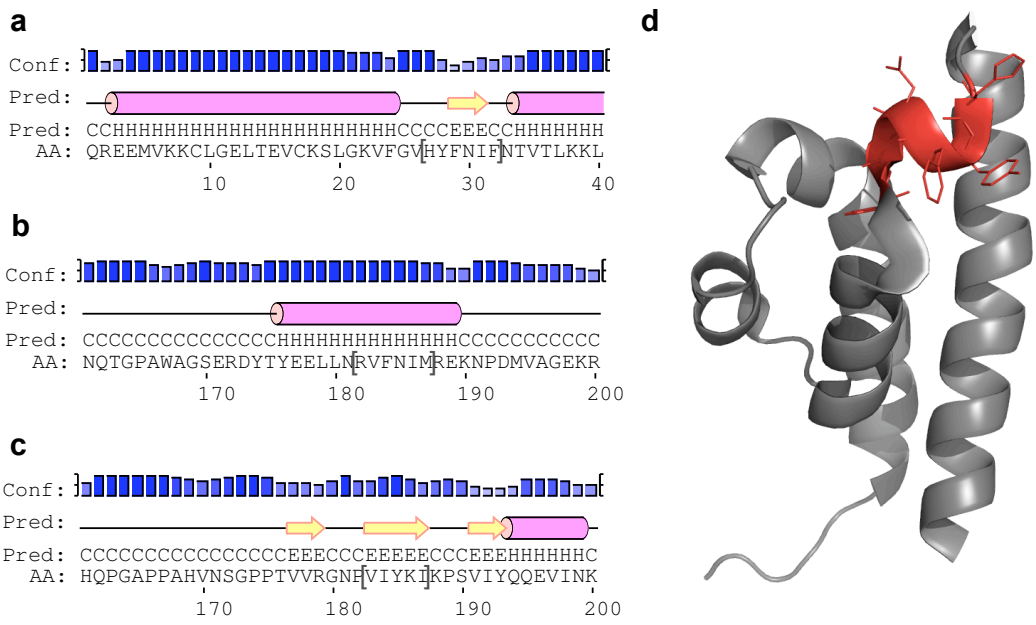


Figure 5.2: The sequence context and predicted native secondary structure of the Waltz peptides HYFNIF (a), RVFNIM (b) and VIYKI (c). α-helices are displayed as cylinders, ß-strands as arrows. The α-helical secondary and tertiary native structure of HYFNIF is shown as red (PDB 2KV2) (Kim and Choi, 2010). Graphics generated in PyMol (DeLano, 2002).

The Waltz sequences taken out of their native sequence and structural context are observed to self-assemble. The following sections describe the structural and biophysical characterisation of these assemblies. Peptides were N-terminally acetylated and C-terminally amidated to avoid assembly effects due to charged termini but also to better represent the possible assembly interactions in a native protein environment.

5.2.2 Fibrillar morphology characterisation

Outside of their native protein sequences the three Waltz peptides chosen for study were found to spontaneously form fibrillar structures upon incubation in water. Transmission electron micrographs reveal that they form amyloid-like fibrillar morphologies, with widths of approximately 20 nm. Each system is morphologically identifiable as amyloid where fibrils are heterogeneous, immeasurably long and unbranched although noticeably thicker in some cases, which can be attributed to increased lateral association. Each system exhibits somewhat unique identifiable characteristic morphologies as shown in Figure 5.3. HYFNIF predominantly adopts

a twisted ribbon morphology with a minority adopting more rope like structures with subtle evidence of twisting. RVFNIM adopts a similar morphology with twisted, where again a minority of RVFNIM fibrils adopt a rope like structure. VIYKI fibrils are observed to adopt almost exclusively a twisted rope like morphology. The morphologies indicate a filamentous hierarchical structure, as observed by the striations within the fibrils. As seen in Figure 5.3 fast Fourier transform filtering has been used to enhance morphological features revealed by the negative staining. These have been used for fibril characterisation as described: HYFNIF exhibits the greatest range of morphologies, which can broadly be described as helical. The simplest helices have a thickest width of 17.04 nm (SD ±1.06 n = 7) and periodicity of 95.16 nm (SD ±3.15 n = 6). More complicated helical arrangements are observed with variable widths and periodicities. After extensive incubation fibrillar structures that may represent a tubular architecture were observed to have widths of 26.03 nm (SD ±3.29 n = 3).

RVFNIM is more consistently found to form tightly wound ‘ropes’ with a width of 17.61 nm (SD ±1.23 n = 5) which have a filamentous structure although the negative stain fails to reproducibly reveal this arrangement. Some examples of filaments twisting into a helical arrangement are also observed with widths at their thickest points of 19.69 nm (SD ±1.59 n = 28). Helical pitch varies dramatically from 147.81 nm (SD ±21.40 n = 11) to immeasurable over a single transmission electron microscopy (TEM) micrograph. VIYKI exhibits the most regular morphology where all observed fibrils form twisted ‘ropes’ with a width of 21.0 nm (SD ±1.2 n = 6) in a helical arrangement with a helical crossover separation of 58.1 nm (SD ±1.7 n = 6) perhaps indicative of helicity (see Appendix I-iii). There is a noticeable decrease in the width of fibres to 17.2 nm (SD ±1.1 n = 2) after extensive incubation. The apparent regularity in morphology suggests a high degree of order.

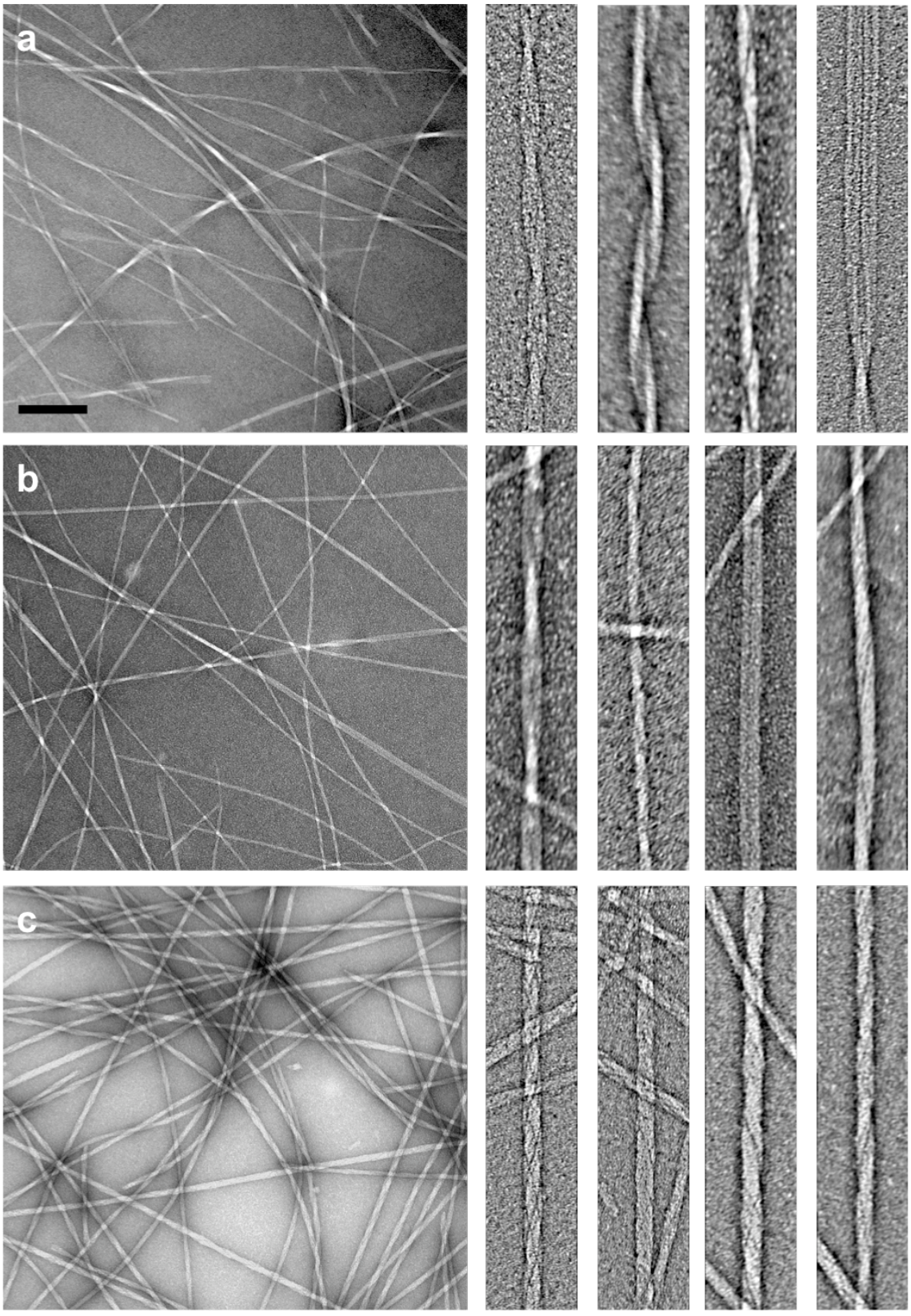


Figure 5.3: The characteristic morphologies of fibrillar HYFNIF (a), RVFNIM (b) and VIYKI (c) as analysed by TEM. The left panels show low magnification where the scale bars represent 200 nm and right panels high magnification images after cropping and fast Fourier transform bandpass filtering – not to scale.

5.2.3 Fibril morphology formation and stability

TEM investigations of the morphology of the Waltz peptides over time post-dissolution reveal immediate fibrillar architecture formation, however, it cannot be ruled out that these peptides were not fibrillar prior to dissolution. If they are monomeric prior to dissolution then these data indicate very fast kinetics of fibril formation for these peptides, although this requires further investigation. Characterisation of morphology over incubation time is more challenging, in particular for HYFNIF and RVFNIM where there is a high degree of morphological polydispersity at any particular time point. Certainly it is clear that their fibrillar structure is stable over extended incubation periods (up to 3 months) as shown in Figure 5.4. Some striking time dependent morphological features that are evident after extensive incubation is that there is a rearrangement of HYFNIF fibrils into structures of 25.95 nm (SD ±1.40 n = 6) nm with a central striation running parallel to the fibre axis. This staining feature may represent two protofilaments laterally associated or stain penetration of a central core consistent with a nanotubular structure (see Figure 3.9). It is tempting to consider these structures as occupying an energetic minimum where they are observed at 72 hours (marked with arrows) and then subsequently predominate at 3 months of incubation. However, although these structures are formed reproducibly they are not always the exclusive morphology at this incubation length (see Appendix I-i) and so this conclusion is suggested tentatively. RVFNIM fibril morphology and morphological polydispersity is maintained over extensive incubation periods. Finally, VIYKI undergoes some maturation through elongation with fibrillar structures being present at 0 hours that are less well-defined and variable in width within the same and between individual fibrils. After 72 hours, twisted rope morphology (see Figure 5.3) is homogenously adopted and stable over extensive incubation.

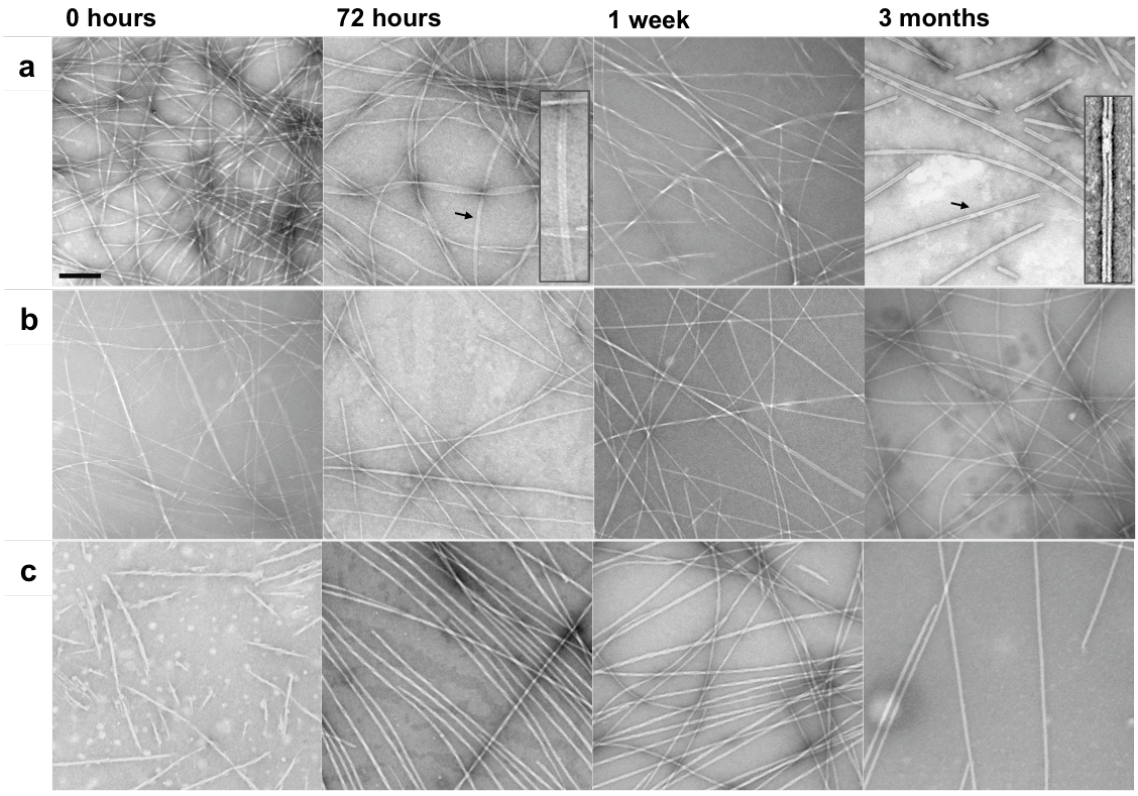


Figure 5.4: Transmission electron micrographs of the Waltz peptides over time. HYFNIF (a), RVFNIM (b) and VIYKI (c) are shown at 0 and 72 hrs to 1 week and 3 months after dissolution. The scale bar represents 200 nm.

5.2.4 Secondary structure determination

The secondary structure of the fibrillar Waltz peptides was investigated using Fourier transform infrared spectroscopy (FTIR) and circular dichroism (CD). Absorbance bands consistent with ß-sheet structure are observed by FTIR with two positive maxima at 1626 – 1633 and 1666 – 1676 cm^{-1}, as shown in Figure 5.5. Although sometimes used to distinguish between parallel and anti-parallel ß-sheets by the presence of the longer wavenumber band at 1695 cm^{-1} (Miyazawa and Blout, 1961, Pelton and McLean, 2000) the ability to do this remains controversial (Barth and Zscherp, 2002) and so the FTIR data can only reliably be attributed to the presence of ß-sheets.

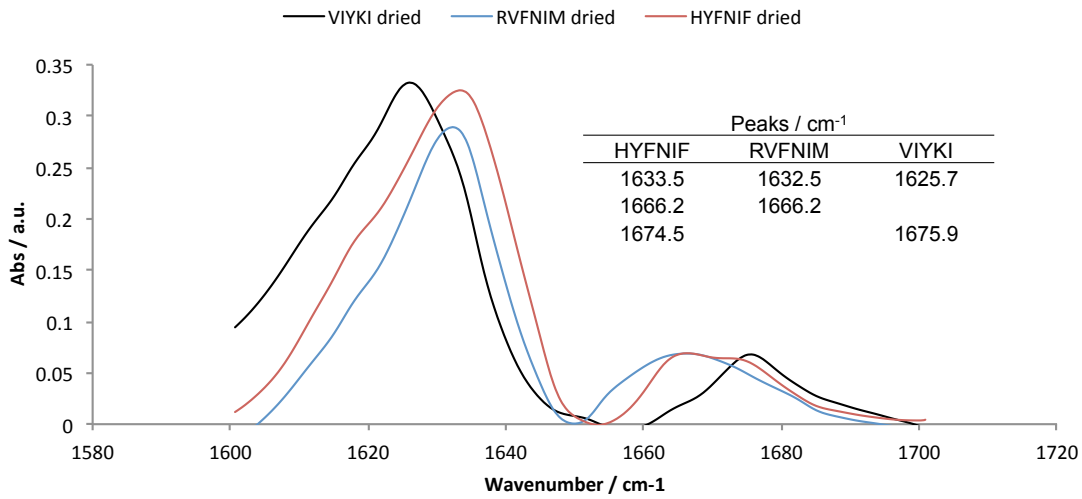


Figure 5.5: The amide I FTIR region of dried films of mature Waltz peptides. Data collected by M. Debulpaep, (Switch LAB, Belgium).

Characterisation by CD was used to corroborate the FTIR data but was found to be complicated by the fibrillar nature of these systems. A classic ß-sheet signal typically has a positive and negative maxima at ~190 and ~220 nm respectively (see Chapter 3) but the signal positions exhibited by the Waltz peptides are not as expected for typical ß-sheet peptides. The CD spectra of the Waltz peptides shown in Figure 5.6 (left panel) show artefacts due to fibrillar anisotropy indicated by the orientation dependence of the signals (see Chapter 3). As well as exhibiting a strong dependence on sample directionality, the signal intensities are between one and two orders of magnitude larger than would be expected for true CD signals. Assuming the poly-L-lysine system is an approximate model for an entirely ß-sheet protein, the CD exhibits an n-π* transition at 217 nm of 18,400 deg cm^2 $dmol^{-1}$ (Townend et al., 1966). These observations taken together can be rationalised to the presence of linear dichroism (LD) artefacts where the fibrillar nature of the sample permits the shear alignment of these systems such that the linearly polarised electronic transitions align and may be seen individually in the CD. Rotation of the sample within the CD reveals which transitions are directionally dependent and shows that for each system the dominant LD artefact is present at ~200 nm in good agreement with the π-π* transition linearly polarised perpendicular to the ß-strand long axis.

To obtain true CD spectra for the Waltz peptides fibrillar samples were sonicated, in an attempt to abolish effects from alignment, as reported in (Andersen et al., 2010). Figure 5.6 (right panel) shows the true CD signals exhibited by the Waltz peptides, after mild sonication. These spectra show little orientation dependence consistent with a shortening of the fibrils preventing the shear forces involved in cuvette loading from aligning them. To ensure the measurements were still representative of structure of the Waltz fibrils TEM was used to confirm the fibrillar nature of the samples as shown in Figure 5.6 (left and right panel insets).

Although confirmed to still be fibrillar the sonication treatment appears to induce random coil CD signals indicative of some fibrillar disruption; these may mask the signals from the Waltz fibrils and as such Dichroweb (URL: http://dichroweb.cryst.bbk.ac.uk/html/home.shtml) analysis was used assess the secondary structure of remaining fibrils in solution. The quantification of the secondary structural elements present in the Waltz fibrils is shown in Table 5.2. The analysis was performed using two data sets and is represented by the standard deviation in each secondary structure percentage value obtained. This analysis indicates that there is indeed a disruption of secondary structure as might be expected considering the treatment of the sample but the remaining structures, confirmed to be fibrillar by TEM, are ß-sheet which is attributed to represent the secondary structure of the Waltz peptides in the fibrillar form.

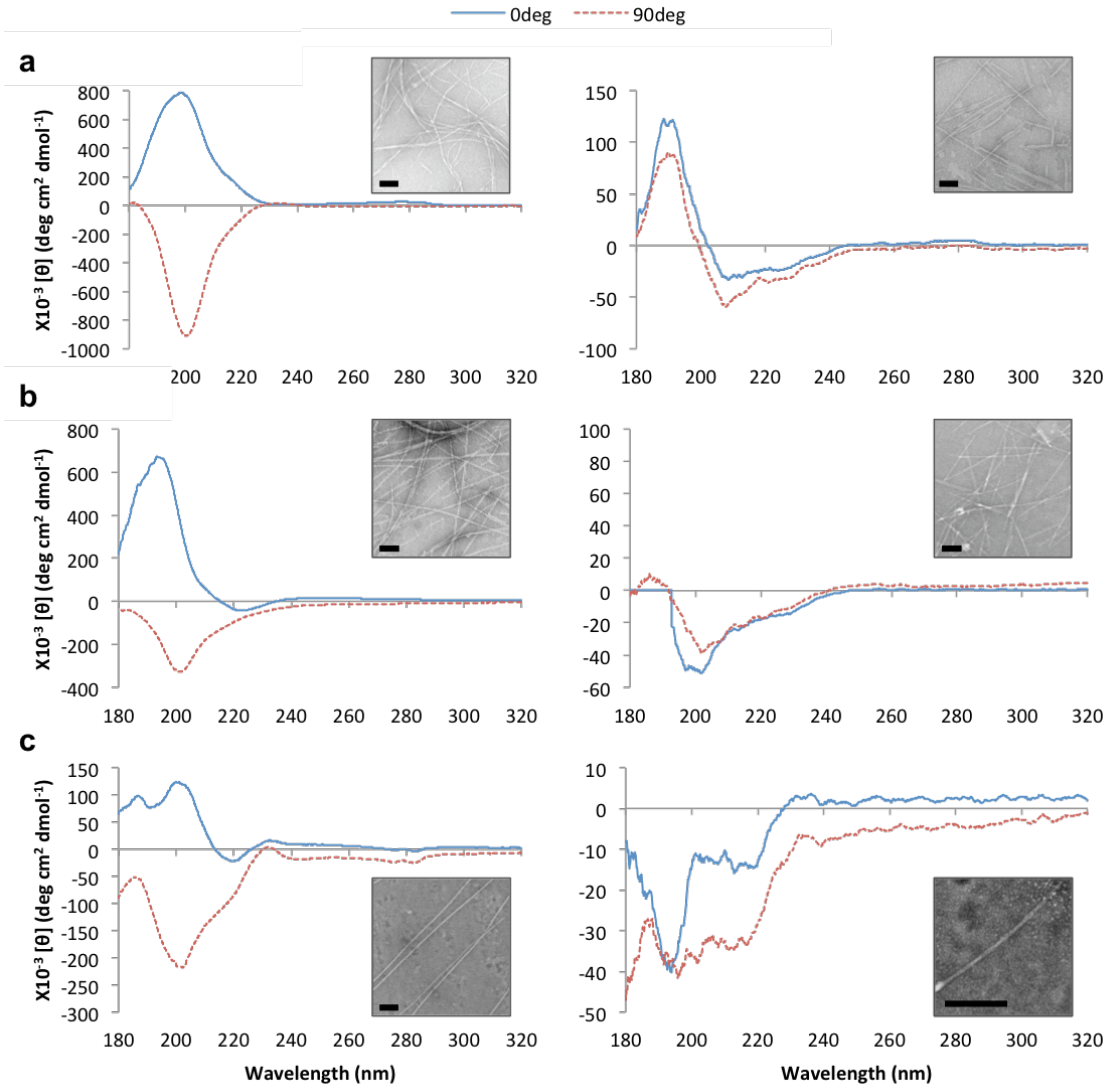


Figure 5.6: CD spectra recorded for HYFNIF (a), RVFNIM (b) and VIYKI (c) pre and post sonication. TEM of the samples are shown as insets, the scale bar represents 100 nm.

The additional artefactual LD signals in the unsonicated samples were deconvoluted (see Appendix I-ii); their presence suggests a high degree of order in the fibrils, and thus presents the opportunity to obtain structural information regarding chromophore arrangement within the fibres. As well as information from the peptide backbone, the unique (RVFNIM and VIYKI) and concurrent occurrence (HYFNIF) of tyrosine and phenylalanine residues in the Waltz peptides also makes these peptides ideal model systems for probing aromatic residue interaction and structure.

2° Structure	HYFNIF		RVFNIM		VIYKI	
	%	+/-	&	+/-	%	+/-
Helix:	6.0	1.4	4.5	2.1	2.0	0.0
Strand:	42.5	3.5	40.0	5.7	42.5	0.7
Turn:	11.5	0.7	12.5	2.1	12.0	0.0
Disordered:	40.0	1.4	42.5	0.7	41.5	0.7
Total:	100.0	7.1	99.5	10.6	98.0	1.4
NRMSD:	0.03	0.01	0.05	0.00	0.04	0.01

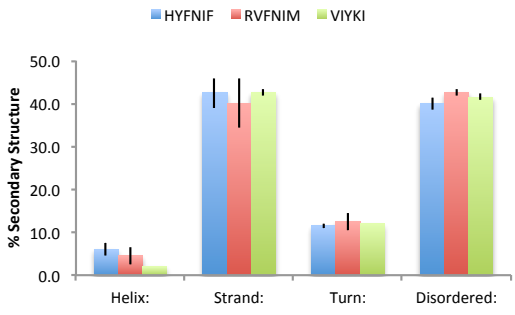


Table 5.2: Dichroweb analysis of the CD spectra recorded for the Waltz peptides post-sonication, highlighting the predominant secondary structural elements.

Where alignment cannot be controlled in the cuvette shear-aligned cases found in the CD experiments using a Couette flow cell fibre alignment was controlled and LD accurately measured.

5.2.5 Secondary structure architecture revealed by linear dichroism

LD on fibrillar systems, in particular amyloid, has been reported previously (Andersen et al., 2010, Bulheller et al., 2009, Marshall et al., 2010). As a measurement of the differential absorption of linearly polarised light parallel and perpendicular to the orientation axis, LD can be deconvoluted to provide structural information on orientated samples. In the Couette cell used in these experiments, the orientation axis is the direction of Couette flow, which is horizontal. Thus the transition moments of the chromophores orientated horizontally and vertically produce positive and negative signals respectively (see Chapter 3) as shown in Equation 3.2.

Using Couette flow to align fibres, any chromophores that are consistently oriented within the sample relative to the fibre axis will produce a signal. Figure 5.7 shows the LD spectra for the Waltz peptides when aligned parallel (=, by Couette flow induced by 1500 – 3000 rpm) and perpendicular (⊥, No Rot) to the orientation axis. Each of the Waltz peptides exhibits a transition at ca. 200 nm which is attributable to the π-π* transition polarised perpendicular to ß-

strand peptides. The sign of the π-π* transition is parallel to the orientation axis with couette flow and perpendicular with shear flow. Where the fibril long axes will be aligned to the orientation axis in Couette flow these data are consistent with the ß-strands of the Waltz fibrils being in a cross-ß arrangement, as also reported elsewhere for amyloid systems (Bulheller et al., 2009).

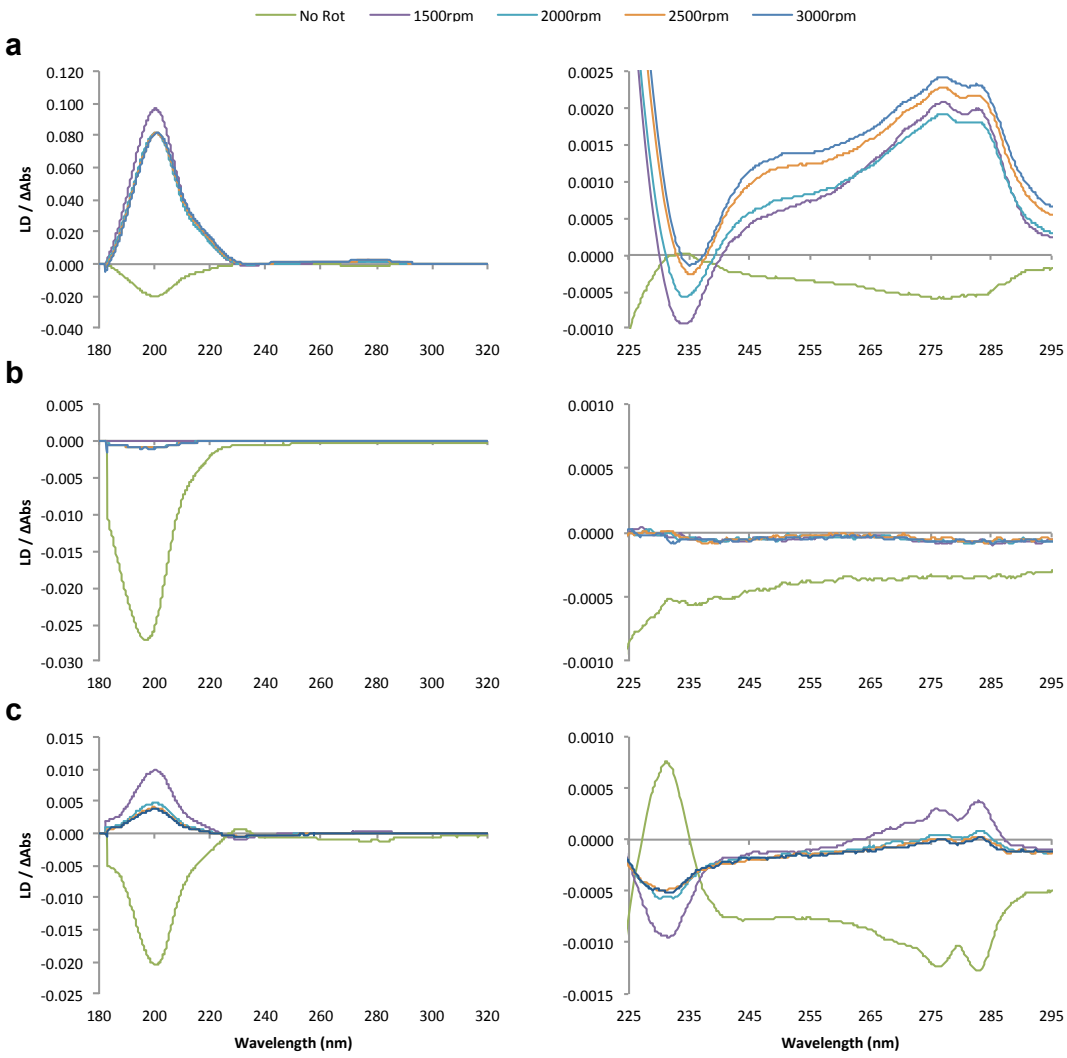
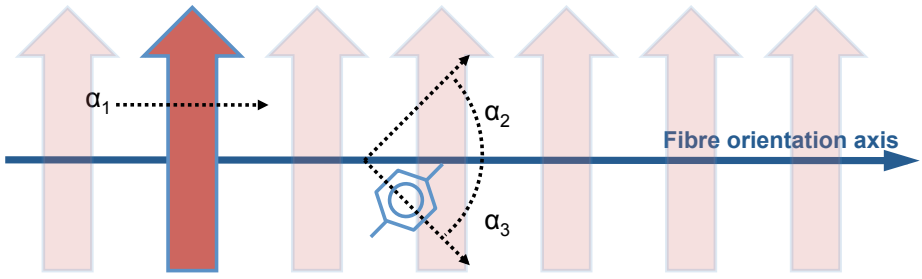


Figure 5.7: The LD spectra arising from shear aligned (No Rot – no Couette rotation) and Couette flow (1500 – 3000 rpm) aligned fibrils of HYFNIF (a), RVFNIM (b) and VIYKI (c).

Signals are also observed in the region expected to arise from aromatic electronic transitions as shown in Figure 5.7 (right panel). The L_a and L_b transitions from tyrosine in the cases of HYFNIF and VIYKI are clearly evident. The L_a transition of tyrosine is parallel to its long axis and occurs at 230 nm while the L_b transition is perpendicular to the long axis and occurs at 275 nm (Bulheller et al., 2009). The single tyrosine present in the VIYKI system giving rise to a signal at 275 nm exhibits splitting due to exciton coupling which can only occur when aromatics are in close proximity (Rodger and Norden, 1997) and this has been previously noted in LD spectra from fibrils formed by the heptapeptide GNNQQNY (Marshall et al., 2010). The tyrosine signal observed for HYFNIF fibrils is also indicative of a high degree of order and regularity in the arrangement of the tyrosine residues within the fibrillar structure. Considered together, this provides strong evidence for aromatic stacking interactions stabilising the HYFNIF and VIYKI fibril systems. Conversely, no signals for phenylalanine are observed for RVFNIM although this could be due to the weaker strength of the transitions associated with phenylalanine. Further, Couette flow also appears to abolish alignment, perhaps suggesting that RVFNIM fibrils are sensitive to mechanical agitation.

Due to the single tyrosine present in VIYKI and the excellently resolved data from this chromophore it is possible to investigate through further analysis the precise atomic orientation of this chromophore within the fibrils. By using the known orientations of aligned chromophores it is possible to calculate the orientations of unknown chromophores using the mathematical descriptions of LD discussed in Chapter 3 as discussed in the following sections.



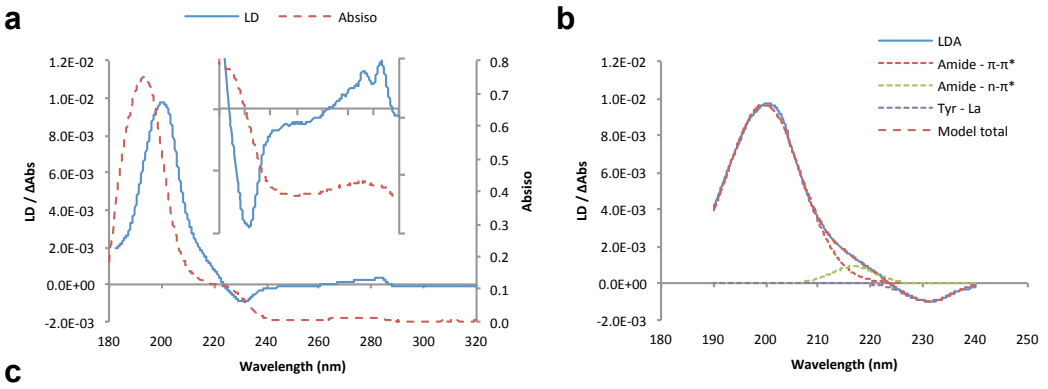
α_1 – 200 nm α_2 – 230 nm α_3 – 275 nm

$$\frac{LD}{A_{iso}} = LD^r = \frac{3}{2}S(3\cos^2\alpha - 1)$$

Figure 5.8: An illustration of the orientation of ß-sheets within the Waltz fibrils, consistent with the cross-ß architecture, based on the LD spectral data. The linear transitions arising from the ß-strands (α_1) are shown in relation to the expected linear transitions of an aromatic tyrosine residue (α_2 & α_3), using these signals the relative angles at which they occur may be calculated.

VIYKI linear dichroism analysis

Figure 5.8 shows the polarisations of the electronic transitions in the Waltz fibrils when assuming the adoption of a cross-ß architecture. By using the equation describing the relationship between LD signal intensity and the angle of the respective transition polarisations (see Equation 3.3 and Figure 5.8) attempts have been made to use the π-π* transition at ~200 nm of the ß-sheets in the cross-ß arrangement to calibrate the calculations required to determine the angle of aromatic chromophore orientations within VIYKI fibrils. Figure 5.9 shows the attempts to this affect, careful analysis was employed to correct for background signal contributions, scattering, and overlapping signals (a & b) such that *S* may be determined and the angles of the polarisations of other electronic transitions calculated based on signal intensity (c). Unfortunately, the calculations were unable to unambiguously determine the chromophore orientations in the VIYKI fibrils; the calculations are not feasible when the angle of the π-π* transition is calibrated at 0 degrees as would be expected for a cross-ß system.



c

Using ß-sheet π-π* as S calibrant

Wavelength / nm	Assignment	$LD^{A}_{r\,/\,decon}$	Known angle	S	α
199.6	Amide - π-π*	1.76E-02	0.00	5.87E-03	0.00
216.6	Amide - n-π*	7.25E-03		5.87E-03	38.78
231.0	Tyr - B_b	-1.44E-02		5.87E-03	#NUM!
276.0	Tyr - L_a	1.90E-02		5.87E-03	#NUM!
283.0	Tyr - L_b	3.23E-02		5.87E-03	#NUM!

Recalibrating ß-sheet π-π* to 33.43 degrees

Wavelength/nm	Assignment	$LD^{A}_{r\,/\,decon}$	Known angle	S	α
199.6	Amide - π-π*	1.76E-02	33.43	1.08E-02	33.43
216.6	Amide - n-π*	7.25E-03		1.08E-02	45.99
231.0	Tyr - B_b	-1.44E-02		1.08E-02	79.10
276.0	Tyr - L_a	1.90E-02		1.08E-02	31.62
283.0	Tyr - L_b	3.23E-02		1.08E-02	0.00

Figure 5.9: The determination of chromophore orientations from LD spectra exhibited by VIYKI. The baselined LD spectra of VIYKI (a) was corrected for scattering and deconvoluted (b) and converted to reduced LD such that the polarisation angles of individual electronic transitions could be calculated (c).

Careful inspection of the reduced LD signals reveals that the transition arising from the L_b state of Tyr in the Couette aligned orientation is in fact more positive than that of the π-π* transition from the ß-sheets (see Figure 5.9). Transitions that are aligned perfectly with the orientation axis are expected to be maximally positive (see Chapter 3) and so the greater intensity of the tyrosine L_b transition would imply that this is more horizontally orientated than the π-π* ß-sheet transition. Based on this assumption, the ß-strands are found to lie at 33.43 degrees relative to the orientation axis (c) and using this the other chromophore angles are calculable. This calculation could have interesting implications for the arrangement of ß-strands within a Waltz fibril. Where the π-π* transition at ~200 nm is considered by the most recent evidence (Rodger et al., 2002) to be polarised perpendicular to the ß-strand direction (i.e. parallel to the C=O bond) two models are consistent with a ß-strand tilt within amyloid fibrils as shown in Figure 5.10.

The fibrils may be perfectly aligned by the Couette flow and ß-strands are tilted within the fibrils, although recent evidence suggests this is not a possible arrangement for cross-ß

structures (Figure 5.10a) (Jahn et al., 2010). Alternatively, the ß-strands may be perfectly perpendicular to the fibril axis but due to protofilament helicity the ß-sheets are tilted between 0 – 33 degrees. At the most extreme case for a fibril of 21.0 nm this would generate a helical pitch of 63.57 nm. Close inspection of the TEM of mature VIYKI fibres finds they are consistent with a twisted three filament model with a tilt angle of 19.85 degrees and a helicity of 174.3 nm (±5.1) (see Appendix I-iii). Using these data it is not possible to ascribe the exact position of the aromatic chromophores and a more complex scenario probably exists than can currently be modelled. However, some interesting questions arise as to whether this helicity explains why the 219 nm transition is not perpendicular to 199 nm in the linear dichroism measurements but also whether it contributes to signal radial averaging in the X-ray fibre diffraction (see Section 5.2.6)?

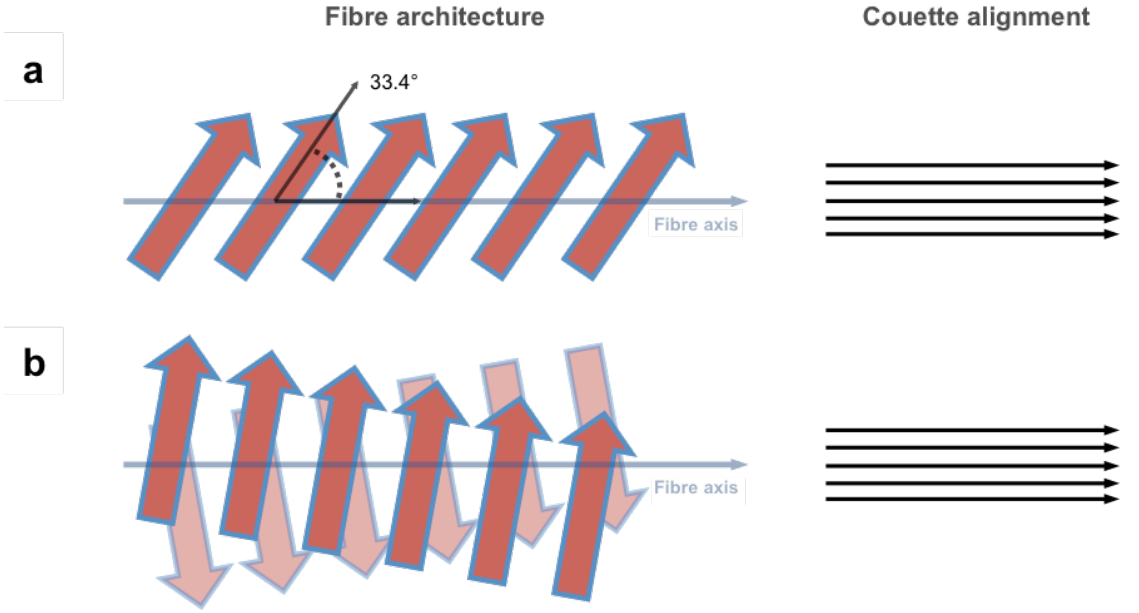


Figure 5.10: The possible arrangements of ß-sheets giving rise to the π-π LD transition. ß-strands may be tilted within a fibre (a) or possibly perpendicular within a helical fibrillar architecture (b).*

Where the TEM data indicate a helical arrangement of protofilaments within the fibrils the second model explaining the apparent ß-strand tilt suggested by the LD data is the most favourable. Due to the discrepancy between the measured helical angle in the TEM and the predicted angle from the LD calculations, the chromophore orientation values have not be used in modelling, yet however the observations here infer interesting molecular and supramolecular features of VIYKI fibrils including helicity and ordered stacked aromatics. To further investigate the arrangement of structural elements within the Waltz fibrils X-ray fibre diffraction (XRFD) was used.

5.2.6 Three dimensional structure revealed by X-ray fibre diffraction

Fibrils of the Waltz peptides were aligned in a variety of ways to ascertain as much structural information as possible, these included alignments in a fibrous, mat and disc texture. XRFD patterns often exhibit broad reflections due to polymorphism and disorder within the alignment, it was hypothesised that a microfocus beam (ca. 5 μm) may sample a smaller region of a fibre alignment, interacting with less of the sample and thus statistically having a lesser chance of diffracting from fibrils with variable structure. This was not found to be the case, however the use of the synchrotron light source granted greater signal to noise ratios, collection to lower resolution (longer length-scales) and so where possible these patterns are presented here.

Fibre alignments

Figure 5.11 shows light microscope images of typical alignments of fibrils of the Waltz peptides. The birefringence evident in each alignment is an indication of a high degree of order within the sample that is reflected in the relatively high quality fibre diffraction patterns shown in Figure 5.12. The XRFD patterns, shown in Figure 5.12, from fibrous alignments of the Waltz peptides show features typical of a cross-ß architecture commonly observed for amyloid (Serpell, 2000) consistent with the spectroscopic data described in Sections 5.2.4 & 5.2.5. The diffraction patterns from the Waltz peptides all exhibit a major meridional reflection at between 4.66 – 4.70

attributed to arise from the spacing between ß-strands along the fibre axis. The patterns differ in the positions and intensities of the equatorial reflections indicating differences in lateral packing arrangements on length scales of up to 40 Å.

The high number of reflections observed in these systems is indicative of a para-crystalline order greater than that typically observed for amyloid systems. The quality of the patterns gives the opportunity to index the reflections and calculate the unit cell parameters within the fibrils. Patterns with a high amount of equatorial information have been collected for other amyloid fibrillar systems before (Makin et al., 2005, Sikorski and Atkins, 2005, Sikorski et al., 2003). In one particular investigation the equatorial reflections arose from lateral hexagonal packing (Papapostolou et al., 2007). In another case fibre diffraction data was found to be consistent with a sub-lattice cell within a larger unit cell (Sikorski et al., 2003). Thus to ensure that the observed equatorial reflections arise from short-range lateral dimensions pertaining to the unit cell and not long range packing arrangements of fibrils as observed in other fibrillar systems, alternative alignments were prepared as shown in Figure 5.13.

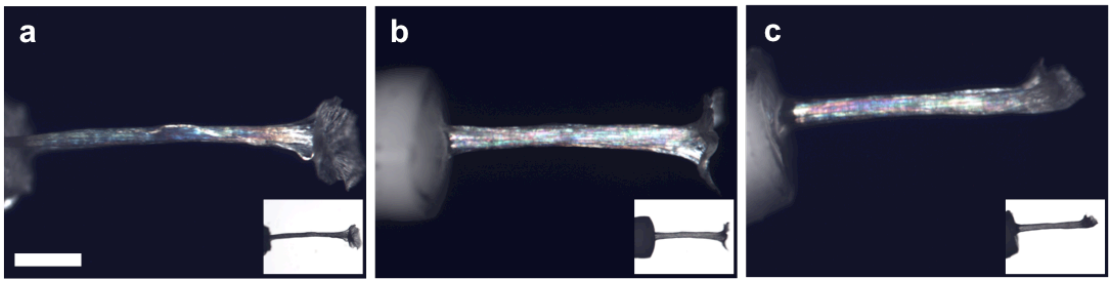
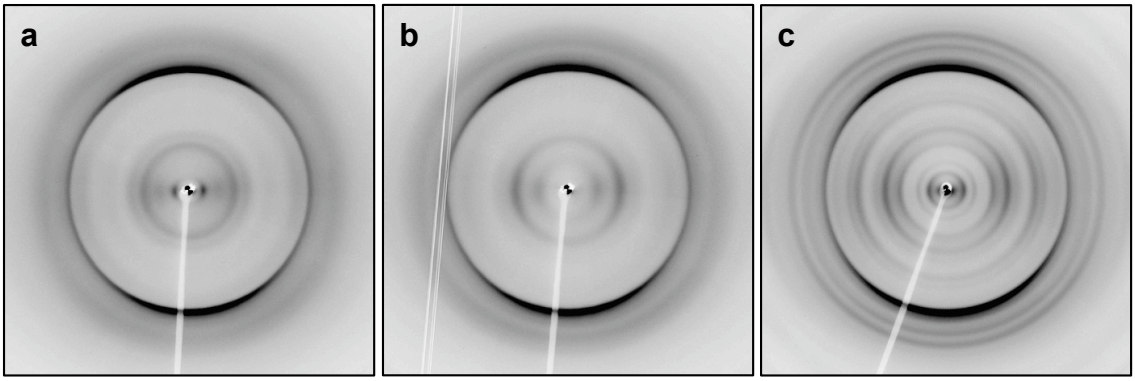


Figure 5.11: Typical alignments of HYFNIF (a), RVFNIM (b) and VIYKI (c) as visualised by cross-polarised microscopy. The insets show the alignments under normal light microscopy. The scale bar represents 500 µm, relevant to each panel.



Equatorials

d-Spacings (Å)	Intensity	+d (Å)	-d (Å)
38.98	233.26	4.19	3.45
19.22	173.73	0.25	0.25
12.37	191.69	0.15	0.14
10.03	156.03	0.12	0.12
7.05	123.81	0.00	0.00
5.23	147.08	0.02	0.02

Equatorials

d-Spacings (Å)	Intensity	+d (Å)	-d (Å)
18.72	180.41	0.23	0.23
10.79	215.09	0.14	0.14
7.36	*	*	*
5.48	144.80	0.06	0.06

Equatorials

d-Spacings (Å)	Intensity	+d (Å)	-d (Å)
33.26	48.83	0.00	0.00
20.04	29.59	0.01	0.01
11.81	39.26	0.02	0.02
9.21	44.28	0.01	0.01
8.09	*	*	*
6.35	37.55	0.01	0.01
5.82	*	*	*
5.12	28.79	0.01	0.01
3.25	22.37	0.01	0.01

Merdionals

d-Spacings (Å)	Intensity	+d (Å)	-d (Å)
4.70	99.95	0.02	0.02
2.38	8.79	0.01	0.01

Merdionals

d-Spacings (Å)	Intensity	+d (Å)	-d (Å)
4.67	129.90	0.01	0.01
2.38	66.65	0.01	0.01

Merdionals

d-Spacings (Å)	Intensity	+d (Å)	-d (Å)
4.66	103.43	0.00	0.00
4.01	35.37	0.00	0.00
2.38	10.74	0.00	0.00

Off-merdionals

d-Spacings (Å)	Intensity	+d (Å)	-d (Å)
3.72	35.54	0.00	0.00

Figure 5.12: X-ray fibre diffraction exhibited from alignments of HYFNIF (a), RVFNIM (b) and VIYKI (c) in a fibrillar texture. All fibre axes are vertical.

Textures

Alternative alignment methods

Where typically fibrils are aligned with all fibril long axes parallel, alternative textures can be prepared such that fibril long axes are aligned on the same plane but are at random orientations to one another. With these sample textures it is a reasonable hypothesis that any preferential packing arrangements should be abolished. As such any reflections arising from long-range order should be absent, the remaining reflections should only be from the lateral unit cell dimensions. The insets of the mat diffraction patterns shown in Figure 5.13 exhibit some preferential alignment of the fibres within a film texture (beam perpendicular to mat plane) and

so cannot completely rule out that there are not contributions from long-range packing to the fibre diffraction signals. Nevertheless, the signals of the diffraction patterns of the alternative alignments match those of the typical fibril alignments and are believed to arise from the repeating cell dimensions. In the case of RVFNIM an additional equatorial reflection is observed, which was included subsequent analysis.

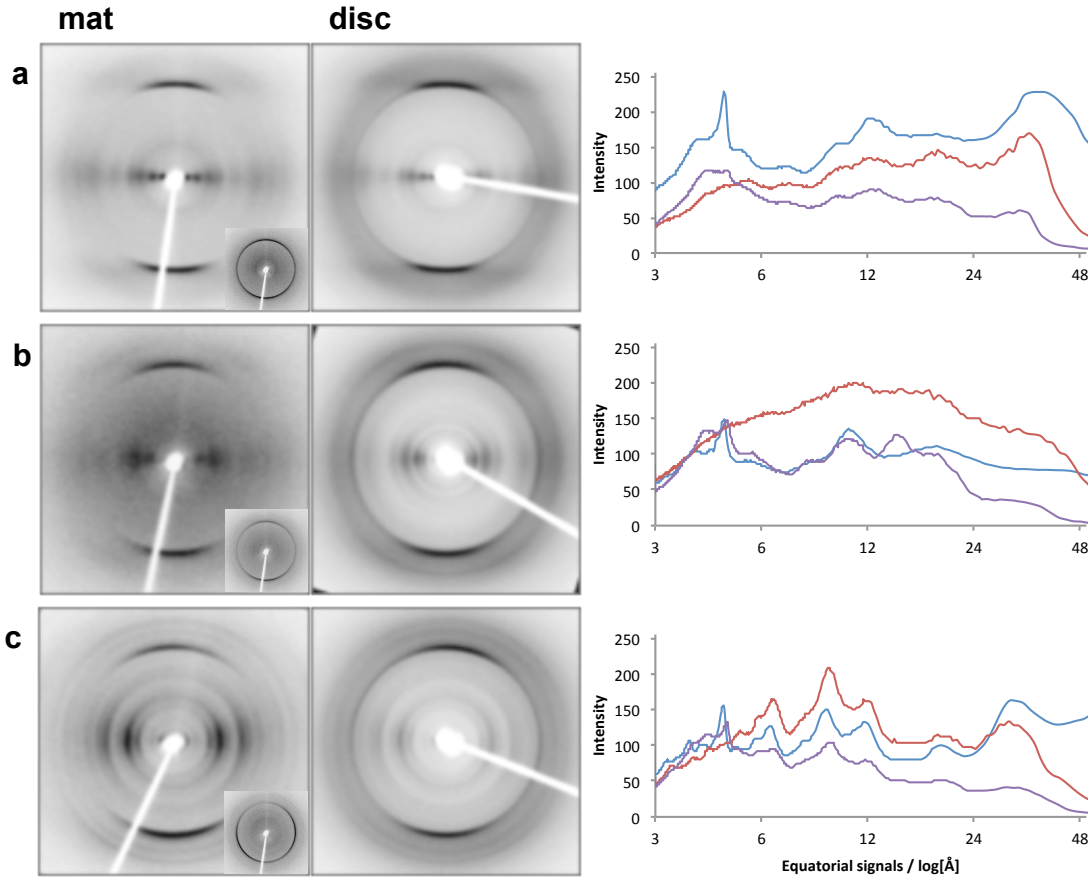


Figure 5.13: X-ray fibre diffraction exhibited from alignments of mature HYFNIF (a), RVFNIM (b) and VIYKI (c) in a film or disc texture. For film textures perpendicular beam shots are shown as insets. Graphical traces of the equators are shown in comparison to those of the fibre alignments shown in Figure 5.12. Alignment textures are shown by colour where blue – fibre, red – film and purple – disc. Patterns are aligned with equivalent fibre axes vertical.

Hydration

To investigate the hydrated state of the Waltz fibrils to tackle the questions of whether the hydrated state structure is the same as the dried but also to further investigate methods of improved alignment a real-time alignment cell was used. Careful consideration was made to the cell functioning in the X-rays as a normal fibre alignment would. The ability to align the cell in the beam was tested using a calibrant prior to diffraction measurements (see Appendix I-iv). Figure 5.14 shows the real time alignment of VIYKI fibrils in an XRFD experiment. The diffuse diffraction ring from amorphous water can be seen at ~3 – 3.5 Å. The meridional and equatorial diffraction signals are shown as graphical traces.

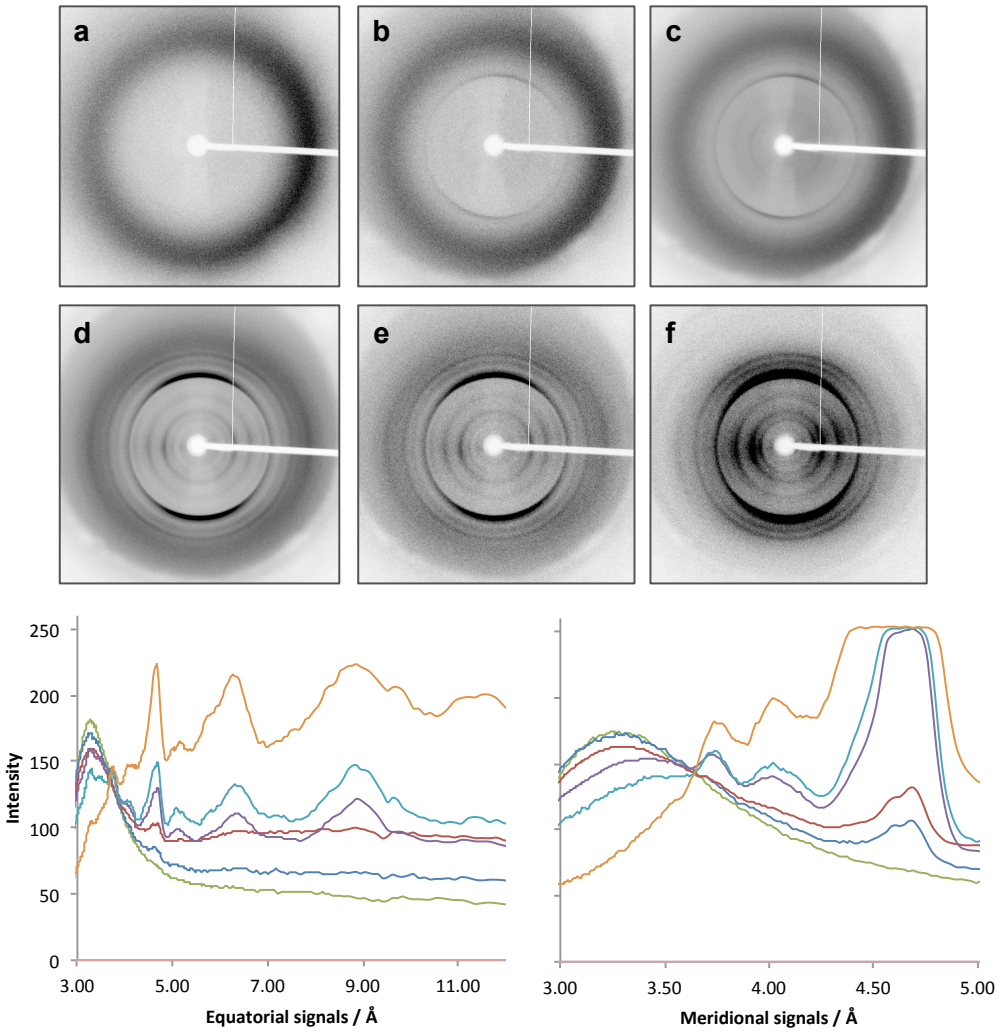


Figure 5.14: Real time alignment of VIYKI at 0 (a), 30 (b), 45 (c), 60 (d), 65 (e) and 70 minutes (f) and graphical representation of the diffraction signals arising from water decreasing as the signals from the fibrils emerge (Green - 0, blue - 30, red - 45, purple - 60, light blue - 65, orange - 70). The contrast ratio between diffraction patterns is preserved for comparison. All fibre axes are vertical.

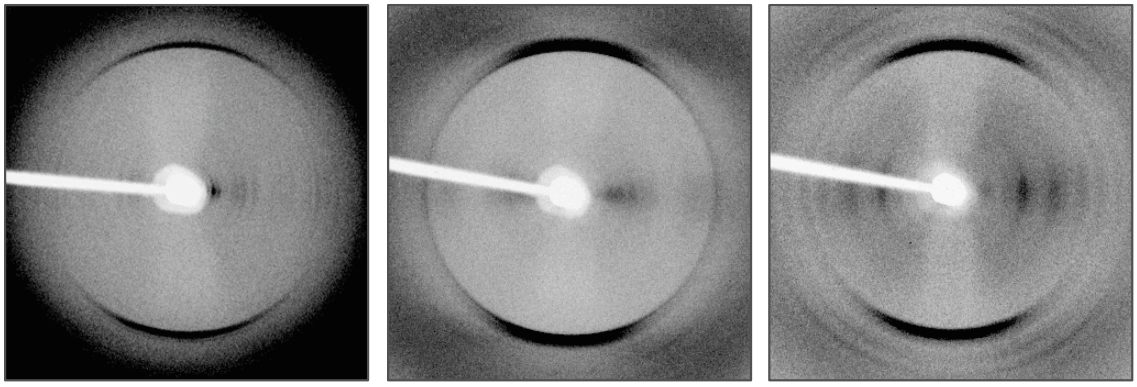


Figure 5.15: X-ray fibre diffraction patterns of the Waltz peptides mature HYFNIF (a), RVFNIM (b) and mature VIYKI (c) in a semi-hydrated state. All fibre axes are vertical.

The water signal predominates the patterns up until 45 mins where it then rapidly decreases in intensity corresponding to solvent evaporation. The diffraction signals arising from the fibrillar peptide then become apparent. Close inspection of the traces, in particular at 60 mins, reveals that the reflection positions and relative intensities are the same as those when dried and the water signal is lost. This strongly indicates that the structure of the fibrils is the same in the hydrated and dried state. Similar data was collected for fibres of HYFNIF and RVFNIM (see Appendix I-v & vi). Also of note is that even in an almost fully hydrated state at 30 minutes the major meridional reflection is present but also this reflection first resolves at 30 – 45 minutes as a doublet. The semi-hydrated fibre diffraction patterns for the Waltz peptides are shown in Figure 5.15, the reflections are generally well matched to the patterns from the fibre alignments further indicating the structures of the fibrils is the same in the hydrated and dried state. Additional equatorial information is observed for HYFNIF but peak positions are not easily measured due to low signal-to-noise ratio and so were not included in subsequent analysis.

Unit cell determination

Using Clearer it was possible to determine the unit cell dimensions (Makin et al., 2007) for the Waltz fibres in an orthorhombic unit cell as shown in Table 5.3. HYFNIF and VIYKI were indexed using the fibre diffraction data. Although alignments of RVFNIM in a film texture exhibited additional equatorial information, the unit cells predicted from these patterns, although similar to those indexed from the fibrous textures, overall were found to be poorer matches; the unit cells indexed from the fibre diffraction data of RVFNIM were thus used. Not only did the use of these unit cells result in better simulated fibre diffraction patterns but also reproduced the additional reflections found in the film texture diffraction pattern. In each case the *a* dimension corresponds to the ß-strand long axis, *b* the inter-ß-sheet distance and *c* the spacing between ß-strands along the fibre axis. Where the meridional reflections arise from a single dimension, it can be assumed that the reflection at ~4.7 Å is the first order reflection corresponding to the *c* unit cell dimension (Morris and Serpell, 2012). Indexing of the unit cells was subsequently performed using the equatorial reflections.

HYFNIF					RVFNIM					VIYKI				
	a	*b*	*c*	Volume / Å^3		*a*	*b*	*c*	Volume / Å^3		*a*	*b*	*c*	Volume / Å^3
Dimensions:	20.86	12.60	4.70	1235	Dimensions:	18.80	21.94	4.67	1926	Dimensions:	34.95	19.01	4.66	3096
	α	*ß*	*γ*			*α*	*ß*	*γ*			*α*	*ß*	*γ*	
Angles:	90	90	90		Angles:	90	90	90		Angles:	90	90	90	
Indexing					Indexing					Indexing				
Observed	h	k	l	Expected	Observed	h	k	l	Expected	Observed	h	k	l	Expected
38.98		N/A		N/A	18.72	1	0	0	18.8	33.26	1	0	0	34.95
19.22	1	0	0	20.86	10.79	0	2	0	10.97	20.04	0	1	0	19.01
12.37	0	1	0	12.6	7.36	0	3	0	7.31	11.81	3	0	0	11.65
10.03	2	0	0	10.43	5.48	0	4	0	5.49	9.21	1	2	0	9.17
7.05	3	0	0	6.95						8.09	4	1	0	7.94
5.23	4	0	0	5.23						6.35	0	3	0	6.34
										5.82	6	0	0	5.82
										5.12	4	3	0	5.13
										3.25	8	4	0	3.22
4.70	0	0	1	4.70	4.67	0	0	1	4.67	4.66	0	0	1	4.66
2.38	0	0	2	2.35	2.38	0	0	2	2.34	4.01	2	2	1	4.07
										2.38	0	0	2	2.33

Table 5.3: The diffraction signals and their respective indexing to the modelled unit cells predicted by Clearer. [0 0 x] indices correlate to the fibre axis, all other are perpendicular to the fibre axis. The unit cell dimensions correspond to the peptides as follows – a – ß-strand long axis, b – ß-sheet spacing and c – ß-sheet hydrogen bonding distance.

5.2.7 Modelling the structural architecture of the Waltz peptides

The unit cell dimensions and lattice determine the positions of the reflections of a diffraction pattern. Figure 5.16 illustrates the equatorial diffraction pattern that arises from the lattice alone (a & ai). The structure within this repeating lattice then modulates the intensity of these reflections such that a new equatorial pattern is observed characteristic of the lattice spacing and constituent structure (b & bi). The intensity profile of the diffracted pattern will be most influenced by the major structural spacings i.e. inter-ß-sheet separation. Importantly though, changes to the side chain conformations can also subtly modulate the equatorial reflection intensities (b & bii).

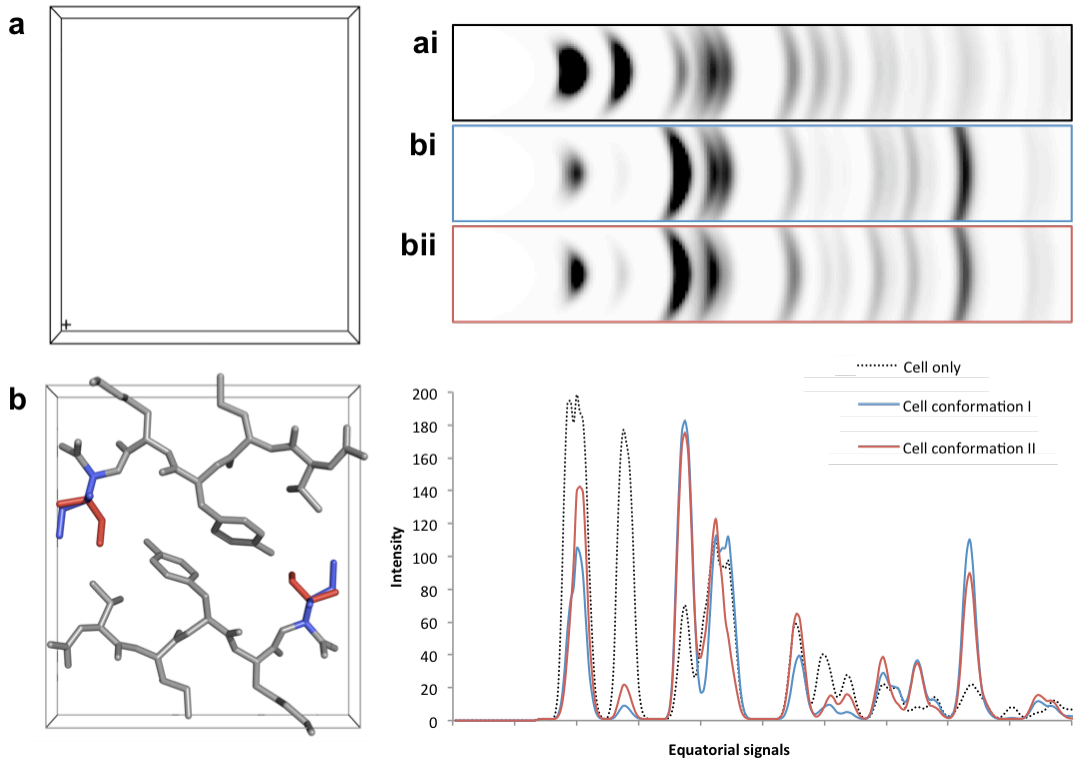


Figure 5.16: Evidence from simulations that fibre diffraction is sensitive to side chain conformation. The repeating lattice without a structure (a) results in an equatorial pattern (ai) whose signal intensities are modulated by the structure it contains (b & bi). The relative intensities of these reflections are thus also sensitive to side chain conformation (b & bii).

On this basis it is possible to explore the unit cell, main chain position and side chain conformation of fibrillar systems based on XRFD simulation. Full simulations from the structures shown in Figure 5.16 are shown in the Appendix I-vii.

Care was taken to ensure models made stereochemical sense by visual inspection, minimisation and molprobity analysis (see Chapter 4). Following this process, models were constructed into the unit cells determined in the previous section, as shown in the Appendix I-viii and further discussed below:

HYFNIF

Unit cell dimensions *- a 20.86 Å, b 12.60 Å, c 4.70 Å, α = β = γ = 90.0 (1235 A^3)*

Peptide dimensions *- x 23.77 Å, y 11.48 Å, z 5.91 Å (843.3 $Å^3$)*

The determined unit cell for this peptide does not predict the low angle 38.98 Å equatorial reflection and no better cell indexings were found. It was concluded that this reflection could represent the diameter of a protofilament. The strongest equatorial reflection of 12.60 Å is attributable to the inter-ß-sheet distance but this is clearly not reconcilable with the low angle reflection and so the 38.98 Å reflection is presumed to arise from the distance of two ß-strands end on end making the protofilament. This is approximately consistent with the dimensions of the peptide modelled in a ß-strand conformation (see Appendix I-viii). In this arrangement the possible packing architectures explored were limited to varying the relative ß-sheet directions and ß-strand displacement parallel and perpendicular to the fibre axis. As such, four architectures were explored for HYFNIF that were based on permutations of the above possible packing arrangements.

RVFNIM

Unit cell dimensions *- a 21.94 Å, b 18.80 Å, c 4.67 Å, α = β = γ = 90.0 (1926 Å^3)*

Peptide dimensions *- x 23.88 Å, y 11.61 Å, z 6.98 Å (779.5 Å^3)*

Based on the dimensions of this peptide, the *a* dimension was found to correspond closely with the ß-strand length and the *b* dimension, with the cell containing two peptides, with an inter-ß-sheet separation of 10.97 Å. In this arrangement the peptide architecture could be explored based on the amyloid architecture class models (Sawaya et al., 2007).

VIYKI

Unit cell dimensions *- a 34.95 Å, b 19.01 Å, c 4.66 Å, α 90.0, β 90.0, γ 90.0 (3096 Å^3)*

Peptide dimensions *- x 18.43 Å, y 11.61 Å, z 6.98 Å (638.3 Å^3)*

Based on the dimensions of this peptide the unit cell likely contains four peptides units. The strongest equatorial reflection of 9.21 Å was assigned as arising from the inter-ß-sheet separation and is approximately half of the *b* dimension and so the unit cell was modelled as being constructed from two filaments of VIYKI consisting of two ß-sheets arranged end on end. Within each filament the arrangements allowed for the amyloid architecture class models (Sawaya et al., 2007) to be explored and so modelling proceeded through a unit cell with half the dimension of *a*.

5.2.8 Structural validation of the Waltz models

Of the ten models constructed for HYFNIF and the sixteen systematically constructed for RVFNIM and VIYKI each model was validated by comparison of its simulated fibre diffraction pattern to its experimental fibre diffraction pattern (see Appendix I-ix). It was found that generally the fully quantitative comparative method detected only very slight differences in the simulated and experimental fibre diffraction patterns indicating the relative reflection intensity changes were small over the difference due to background scattering. Methods were employed to minimise this effect (see Chapter 4) and in addition using a visual qualitative and a fully

quantitative comparative method unlikely architectures were first ruled out to identify the best possible models that represent the fibrillar structure of the Waltz peptides. The fully quantitative comparison method in some instances identified models as good matches which by visual inspection were poor matches but using this method it enabled the selection of the most representative models with consideration of the various simulation comparison methods. Figure 5.17 shows the simulated XRFD for the models constructed that best represent the fibrillar structure of the fibrillar Waltz peptides.

Subsequently, ß-sheet displacement parallel to the fibre axis was found to be important for meridional reflection matching in RVFNIM and VIYKI. In RVFNIM, and to an extent with VIYKI, a clear relationship between ß-strand displacement and the matching of the first major meridional reflection position was found. In this case the second order weak reflection at ~2.38 Å was found to always be half the distance of the unit cell *c* dimension. However the first strongest meridional reflection position was dependent on whether the peptide was in a register or parallel displaced arrangement.

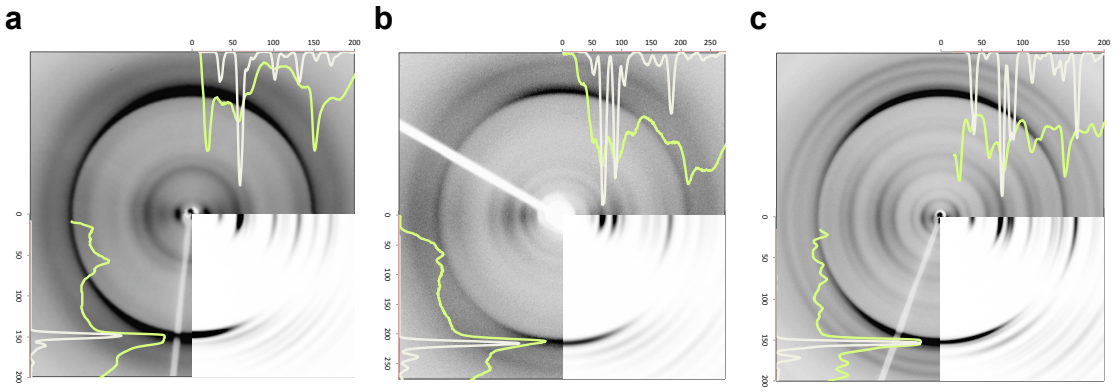


Figure 5.17: A comparison of the experimental and simulated X-ray fibre diffraction exhibited by the Waltz peptides HYFNIF (a), RVFNIM (b) and VIYKI (c). The simulated patterns shown as insets represent the best matches produced by the modelled structures. An overlaid graphical trace compares the equatorial and meridional intensity of the experimental (green) and simulated (white) X-ray fibre diffraction signals. Traces have been scaled for comparison. All fibre axes are vertical.

Figure 5.18 shows the relationship between ß-sheet displacement and the meridional reflections exhibited. In the case of RVFNIM with an indexed *c* dimension of 4.67 Å having a miller index of [0 0 1], the [0 0 2] reflection will occur at 2.34 Å. However the second order meridional of the pattern occurs at 2.38 Å, the first order of which would be 4.76 Å in disagreement with the major meridional reflection in the experimental pattern at 4.67 Å. ß-strand displacement with a *c* cell dimension of 4.76 Å modulates the intensity of the [0 1 1] reflection at 4.67 Å, showing that the pattern of meridional signals at [0 1 1] and [0 0 2] is indicative of the ß-sheet parallel displacement/out of register arrangement. A similar relationship was found for VIYKI and although the same phenomenon was not identified for HYFNIF the parallel displacement of ß-strands end-on-end is different to the displacement of ß-sheets face-to-face perhaps changing the behaviour of meridional diffraction signals in this system.

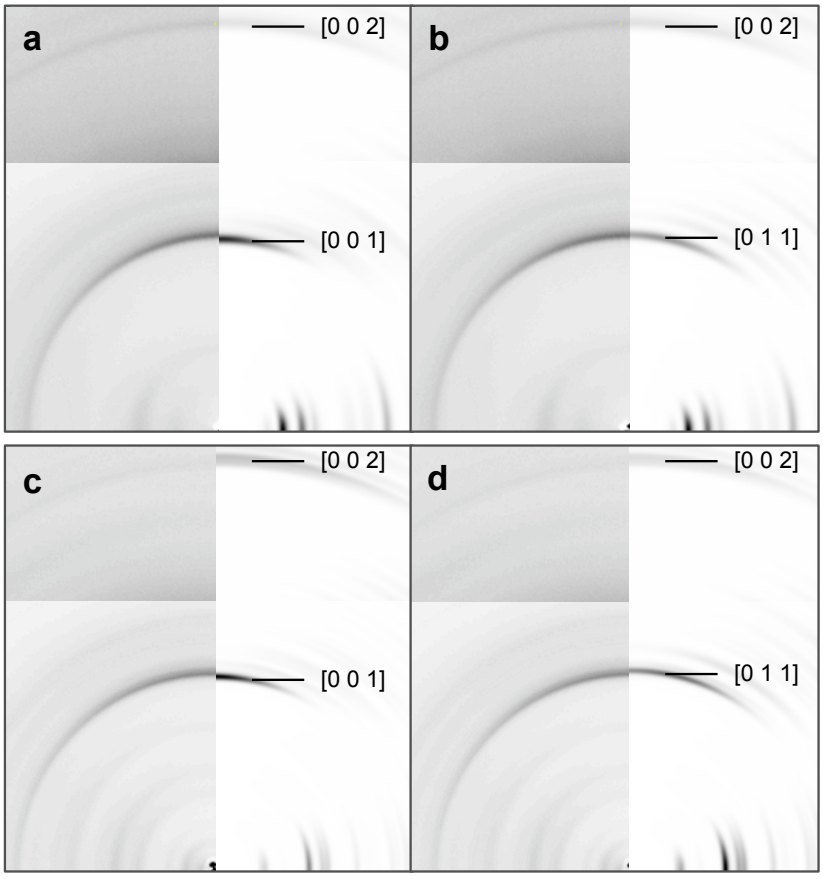


Figure 5.18: A meridional comparison of in register and parallel displaced Waltz models including class II RVFNIM in register (a) and out of register with class IV VIYKI in register (c) and out of register (d).

Figure 5.19 shows the best models corresponding to the simulated patterns shown. It was found that other architectures to an extent also fit the available data but the models shown here represent the best matches constructed. This raises questions over whether the fibrils may contain mixed architectures or that architectures differ between fibrils within a sample, as is the case for the commonly encountered heterogeneity of amyloid fibrils. Each model is, as expected, stabilised by interamide backbone hydrogen bonding though each peptide adopts a differing lateral packing arrangement. HYFNIF was found to adopt a perpendicular ß-strand displaced architecture. RVFNIM adopts a class II in register parallel arrangement where ß-sheets are face-to-back and additional stabilising interactions include hydrogen bonding between lysine and asparagine of adjacent ß-strands. VIYKI adopts a class I out of register antiparallel face-to-face arrangement with a very tight interdigitation of side chains as also observed for recent crystal structures of amyloid-like peptides (Nelson et al., 2005, Sawaya et al., 2007).

It is interesting that this data indicates that the Waltz algorithm does not predict a particular architecture of amyloid. This is noteworthy when one considers the sequence similarity of the three peptides reported here: that is the common aromatic moiety at position three, the semi conserved residues asparagine or lysine at position four and the totally conserved isoleucine at position five. It might have been expected to find common rules and interactions that define the assemblies of the Waltz peptides identified but instead based on the model structures proposed an adaptability in the architectures formed is observed. The common theme that is apparent between these models is the tight packing of peptide molecules. Although an additional hydrogen bonding interaction is observed in the RVFNIM model the main stabilising factors appear to be the characteristic interamide backbone hydrogen bonding and favourable steric contacts.

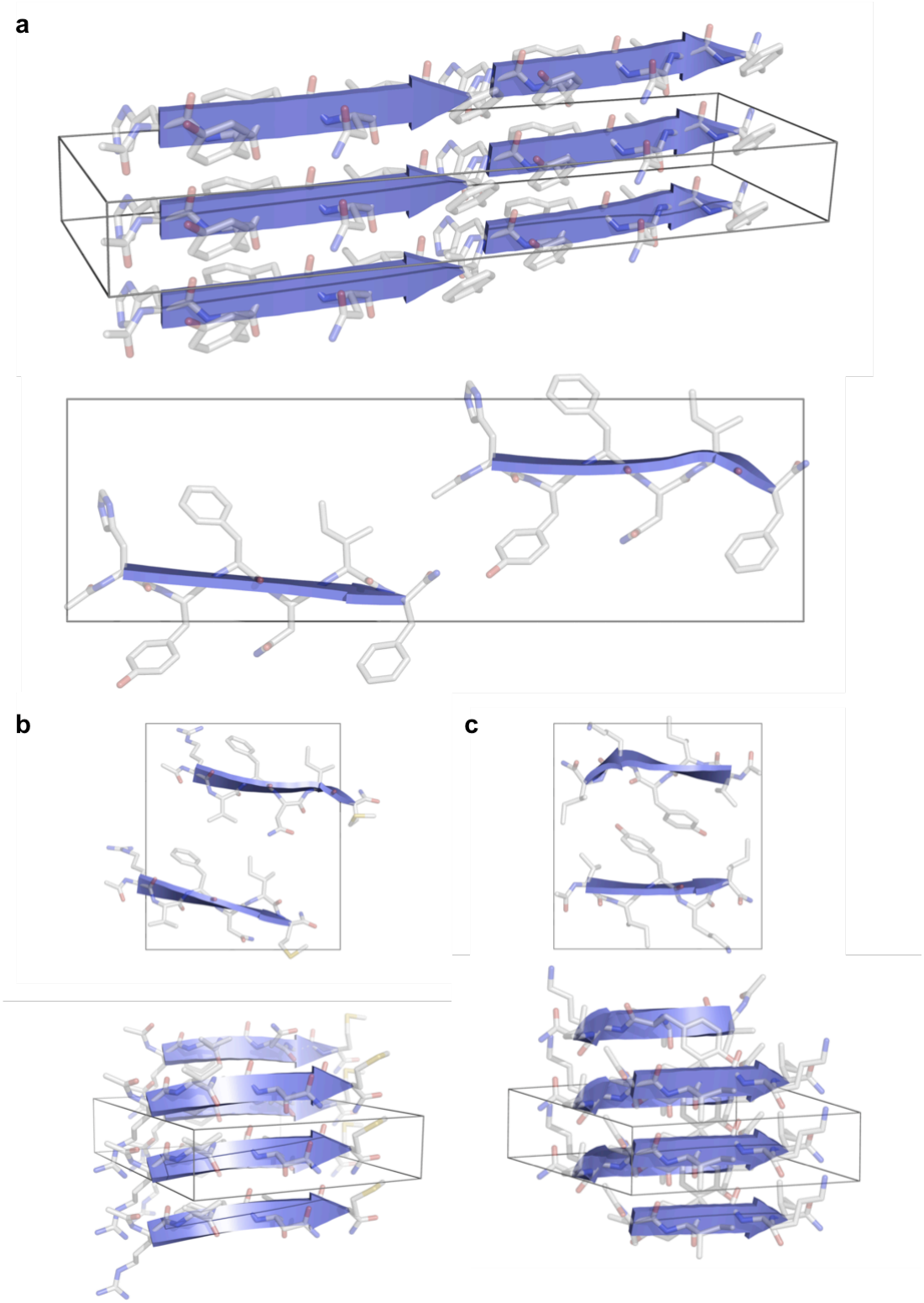


Figure 5.19: The most representative models of the Waltz peptides in the fibrillar form; HYFNIF (a), RVFNIM (b) and VIYKI (c). Graphics generated in PyMol (DeLano, 2002).

5.3 Summary

This work further confirms the Waltz algorithm's ability to predict ordered aggregation, characterised here are two hexapeptides peptides and one pentapeptide uniquely identified during the iterative sequence exploration that developed Waltz (Maurer-Stroh et al., 2010). Each Waltz peptide adopt a cross-ß architecture but with varying lateral packing arrangements. The sequences are found to have different ß-sheet propensities but in their native structures adopt a conformation dependent on their surrounding amino acid sequence, removed, the Waltz peptides adopt the cross-ß arrangement. On the nanoscopic scale these systems are observed to adopt characteristic fibril morphologies that may be based on the underlying differing architecture but are ultimately typical for amyloid fibrils (Goldsbury et al., 1997). The exact relationship between these two levels of structure would be fascinating to investigate, likely requiring the utilisation of cryo-EM to probe this hierarchy. The investigations conducted on these peptides has required careful consideration of the use of the biophysical and structural techniques employed. It was found that circular dichroism data is highly dependent on the fibrillar nature of the assemblies but highlights methods, also reported elsewhere (Andersen et al., 2010, Dafforn et al., 2004, McDonald et al., 2008), for the successful interpretation of this data. The identification of LD artefacts in CD data has been reported before (Norden, 1977). These data clearly demonstrate that LD artefacts can occur in CD experiments, the identification of which here having broad methodological implications for using CD to study systems exhibiting anisotropy.

Clearly care should be taken in analyses of these sorts of data but the effects can be usefully rationalised and used in linear dichroism experiments to reveal information about not only chromophore orientation but possible insights the helical supramolecular arrangement of fibrils within an amyloid-like fibril. Particular care was taken to ensure that the hydrated state of these systems was the same as the dried through the use of a humidity chamber, the structures here are found to be unaffected by drying as reported elsewhere (Squires et al., 2006).

Using Clearer, unit cells were determined for the Waltz peptides and the specific architectures that best represents the fibril structure of the Waltz peptides. Nevertheless, judged by qualitative and quantitative analysis it is possible that the actual fibrils contain a mixture of these architectures where the difference between simulated patterns is slight. This is particularly apparent for RVFNIM but it interesting where other biophysical techniques imply that the structures of RVFNIM are more disordered or perhaps indeed more heterogeneous than HYFNIF and VIYKI. It would be tempting to use the exhibition of tyrosine exciton coupling in the LD measurements to further justify the choice of the VIYKI architecture, however the same exciton coupling is observed for HYFNIF though the model does not have the same tyrosine proximity. It has been noted before that this phenomena cannot distinguish between tyrosine exciton coupling between adjacent ß-sheets or between stacked ß-strands (Marshall et al., 2010). Continued investigations, however, are demonstrating that di-tyrosine formation is more efficient in VIYKI than HYFNIF (data not shown). Although the same distinction must still be made regarding tyrosine interactions between adjacent ß-sheets or stacked ß-strands, these data seems to support the models chosen here.

Structurally, RVFNIM and VIYKI are found to adopt an out of register arrangement by parallel displacement of ß-sheets with respect to one another. This has been observed for crystallised short amyloidogenic peptides (Sawaya et al., 2007) but at the time of writing the data here is the first example of this being demonstrated through experimental and simulated XRFD. The arrangements adopted are variable between the Waltz peptides, which is perhaps surprising given the position sequence similarity of these peptides but is a reflection of the adaptability of this conformation. Consideration should be given to mutagenesis studies on these peptides within the confines of the residues the Waltz algorithm predicts can occur at each position however consideration should be made over whether changing a residue will be a modification or, given the fact that the system is so small, create a completely new system.

In future studies of these systems, in particular VIYKI, will make valuable unique and additional contributions to the model systems currently describing the structural space of amyloid assemblies. The characterisation of these systems not only further confirms the Waltz algorithms ability to predict ordered amyloid aggregation, but also introduces an opportunity to further understand this conformation whilst also presenting new highly ordered and well characterised systems for nanotechnological applications.

Waltz's design was such that it inherently would not overpredict hydrophobic amyloidogenic sequences and underpredict polar sequences, although the peptides presented here represent largely a hydrophobic sequence space for amyloid investigations will continue to characterise other sequences predicted by Waltz.

6 Results and Discussion:

A fragment of α-synuclein adopts a novel nanotubular cross-ß architecture

6.1 Introduction

α-Synuclein (α-Syn) is associated with the development of neurodegenerative Parkinson's disease (Goedert, 2001). Attributable to the disease's symptoms, patients exhibit degeneration of the *substantia nigra* resulting in dopamine deficiency but also the concurrent deposition of Lewy bodies (Forno, 1996). The unambiguous identification of α-Syn in the Lewy bodies of Parkinson patients (Spillantini et al., 1997) identifies a causal link between α-Syn and PD. α-Syn is a 140 amino acid polypeptide (as shown in Figure 6.1) and has long been thought to be a natively unfolded monomeric protein (Weinreb et al., 1996) that adopts an α-helical structure upon vesicle binding in its apparent native functional role (Davidson et al., 1998). Recent evidence suggests that under physiological conditions α-Syn has an α-helical tetrameric structure, the destabilisation of which must first occur to lead to misfolding, aggregation and ultimately the pathology of Parkinson's disease (Bartels et al., 2011). Assemblies of α-Syn have been observed for synthetically formed fibrils of this protein *in vitro* and have been confirmed to have a ß-sheet rich structure consistent with the cross-ß architecture (Serpell et al., 2000a).

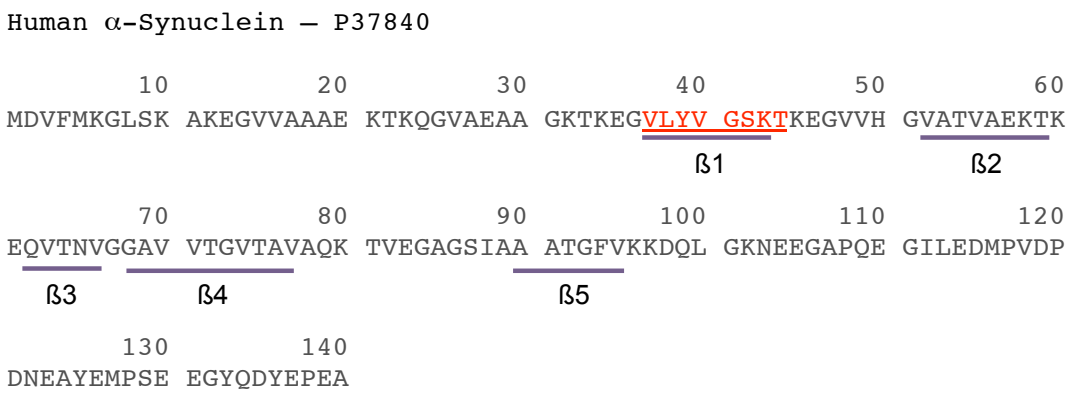


Figure 6.1: The sequence of α-synuclein and the αSß1 sequence highlighted in the context of the ß-forming segments identified in (Vilar et al., 2008).

Solid state and hydrogen/deuterium exchange NMR studies by Vilar *et al.* indicate that the amyloid architecture of full-length α-Syn comprises of five ß-strands (ß1 – 5) in a cross-ß arrangement creating a series of consecutive ß-hairpins. These stack to create five parallel ß-sheets extending along the direction of the fibre axis (Vilar et al., 2008). The rational and current literature regarding the use of short amyloidogenic sequences have already been discussed in detail (see Chapter 1 and 5). Reported here using a ß-sheet propensity algorithm called SALSA (simple algorithm for sliding averages) (Zibaee et al., 2007), core amyloidogenic regions of full-length α-Syn were identified with the aim to add more detail to the current structural understanding of α-Syn. One such fragment, 37-N-VLYVGSKT-C-44 (herein referred to as αSß1), was found at high concentration to form elaborate fibrillar structures consisting of a cross-ß architecture that adopts a long-range supramolecular nanotubular organisation. Although nanotube morphologies have been reported before, at the time of writing, observations of this morphology have not been made on the same length scales and with a linear peptidic system. There is an increasing interest in supramolecular assemblies and the architectures they form, with a wide range of conceived applications (Valery et al., 2011). Nanotubular morphologies not only present an enticing challenge to be able to understand, but offer opportunities to develop the methods available to probe molecular packing in and the interactions that stabilise self-assembled structures.

Dr. S. Zibaee (MRC, Cambridge) is acknowledged for assembling and providing the α-Syn fragment. Dr. P. Sikorski (NTNU, Norway) is gratefully acknowledged for his collaborative efforts in indexing the unit cell used in the final modelling.

6.2 Results and Discussion

6.2.1 αSß1 assembled morphology by transmission electron microscopy

The αSß1 peptide dissolved in water at high concentration was found to form a gel. At low concentrations, structures were not observed in transmission electron microscopy (TEM); though as concentration was increased, semi-transparent tube-like structures were observed that appear to consist of twisted flat ribbons (Dr. Shahin Zibaee, verbal communication). Figure 6.2 shows that the αSß1 structures vary in width from ~240 (i) to ~335 nm (ii) where the thinner structure extends to reveal a helical tape unravelling from a tube-like structure. The twisted ribbon has a width of ~363 nm (iii), suggesting that the larger structures may in fact be tapes that have entirely unravelled from the tube-like structures. The unravelled tapes may be compared to 'nanobelts' reported by other investigators (Cui et al., 2009). Close inspection of the inset of Figure 6.2a reveals that the flat ribbon splits into two (iv) before re-joining to form the original tape structure. This suggests that the interactions stabilising the tapes perpendicular to the tape long axis are non-covalent, i.e. hydrogen bonding or Van Der Waals, such that they can be overcome but then may reform. Figure 6.2b again clearly shows that the flat ribbons when unravelled may vary in their width considerably. The helical pitch varies and due to the transparency of the ribbons it is difficult to assign left or right-handedness. The inset of Figure 6.2b reveals the close association of three helical tapes constituting the mature tube from which they extend.

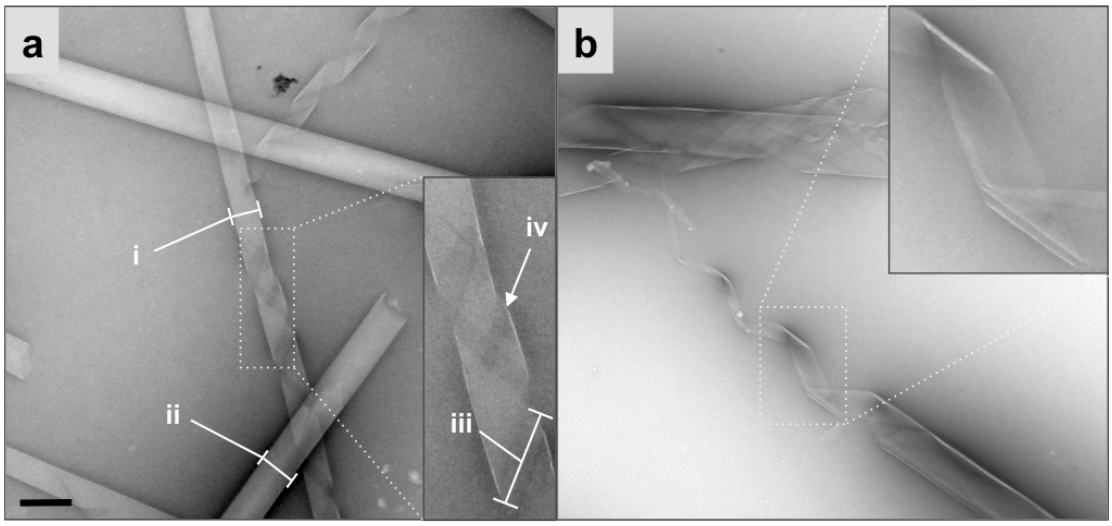


Figure 6.2: The morphology of αSß1 structures as visualised by TEM. The scale bar represents 500 nm.

Comparison can be made to other studies of nanotubular self-assemblies finding morphological similarities to systems that bear no relation to the αSß1 peptide. In particular, characterisation of nanotubes formed by the self-assembly of 1,4,5,8-naphthalene-tetracarboxylic acid diimide (NDI)-lysine amphiphiles is useful for the classification of the structures observed here (Shao et al., 2011). Figure 6.3 shows the variable morphology of αSß1, again highlighting the tendency of the tapes to laterally separate but once again reform the original tape structure (a). By the definitions of Shao *et al.* coiled ribbons (b) and helical tapes (c) are also observed (Shao et al., 2011) who attribute these structures to early intermediates on the pathway to the assembly of mature tubes as observed in Figure 6.2.

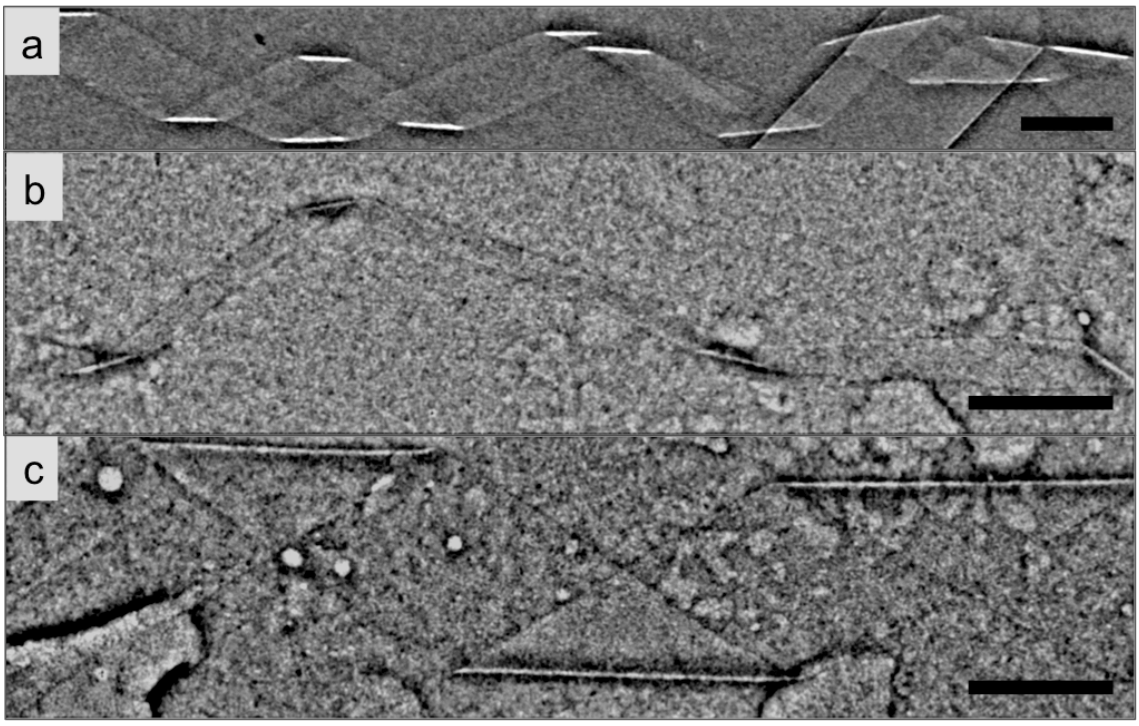


Figure 6.3: The variable morphology of αSß1 visualised by TEM. Micrographs have been processed by FFT bandpass filtering. Scale bars represent 200 nm.

To ascertain how the αSß1 peptide is arranged in these nanotubes, fibrils were aligned using different methods to prepare a variety of textures for X-ray fibre diffraction (XRFD) analysis.

6.2.2 αSß1 assembled structure determination by X-ray fibre diffraction

X-ray fibre diffraction patterns obtained from aligned fibres of αSß1 share similarities with the characteristic cross-ß pattern observed for amyloid fibres with major meridional reflections at 4.75 – 4.81 Å but a larger major equatorial spacing of 27.13 – 29.69 Å. The pattern differs from classical cross-ß fibre diffraction where the number of meridional reflections outnumber the equatorial reflections suggesting that the complexity in this structure occurs along the fibre axis and but not perpendicular to it, as it is normally found for amyloid systems. Further, the presence of off-axis reflections implies a helicity in these structures, perhaps occurring along the fibre axis - as would also be consistent with TEM observations.

Fibrils of αSß1 were prepared in three textures as shown in Figure 6.4 (see also Table 6.1) - fibrous, mat and disc. The latter two textures should be equivalent and may both be described as films. The diffraction exhibited by aligned fibrils of αSß1 suggested a complex fibrous texture where the 9.95 Å (± 0.03) reflection axis changes from off-axis with 6^{th} order rotational symmetry (θ – 0°) to meridional (θ – 90°) upon rotation about the fibre axis (see Figure 6.4a). In an attempt to obtain more structural information about the system and understand this complex texture, alternative alignments were prepared as described previously (Makin and Serpell, 2005b). The mat texture was found to have some preferential alignment (see Figure 6.4b) and surprisingly was found to have a texture comparable to the fibrous alignment. With the X-ray beam parallel to the plane of the mat a pattern is exhibited that almost entirely reproduces the θ – 90° fibrous pattern as shown in Figure 6.4a. Additional equatorial reflections are also observed. In the disc texture, except for the poor resolution of the equatorial reflections at 19.49 Å (+0.87/-0.80) and 13.39 Å (+0.14/-0.14), the most information is revealed; reflections are sharp and confined to discrete axes and there appear to be layer lines exhibited. The alignment was found to be fully radially averaged about the plane of the disc (data not shown) but exhibited diffraction comparable to the fibrous and mat textures only with better resolution. It would seem that regardless of the texture adopted that patterns report the same structure and thus can be used in a combined analysis of the reflections.

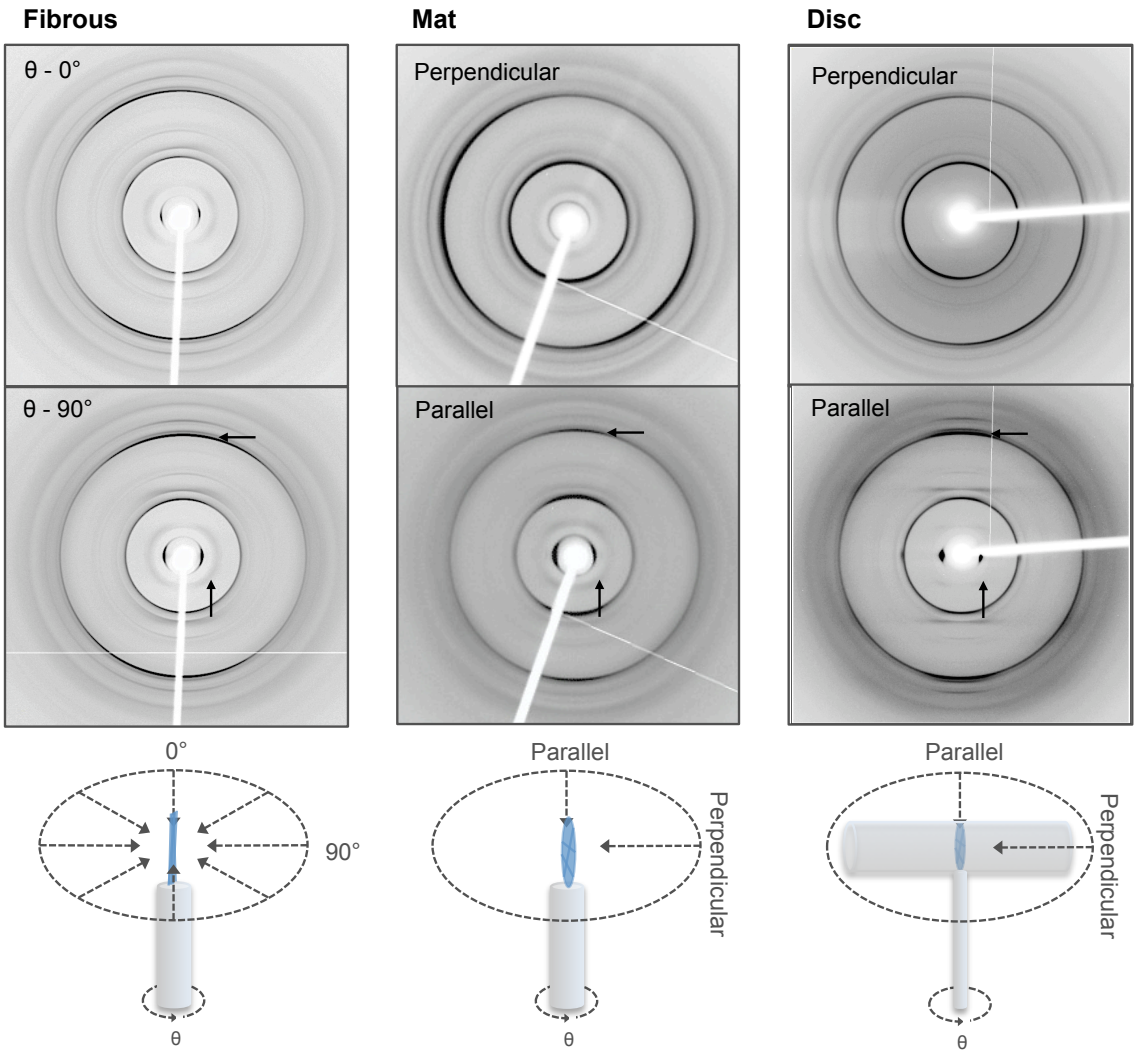


Figure 6.4: X-ray fibre diffraction from a variety of alignment textures of fibrils of aSβ1 including fibrous, mat and disc. Schematics of the alignments in the X-ray beam are also shown. Parallel and perpendicular terminology refers to the beam position relative to the plane of the film. All fibre axes and meridians are vertical. Key off axis ~17 Å reflections are labelled with vertical arrows and the ~4.7 Å meridional reflections are labelled with horizontal arrows.

Fibrous – 90°

Equatorials

d-Spacing (Å)	Intensity	+d (Å)	-d (Å)
29.60	216.97	0.09	0.09
-	-	-	-
-	-	-	-

Merdionals

d-Spacing (Å)	Intensity	+d (Å)	-d (Å)
9.95	211.13	0.03	0.03
8.58	76.36	0.05	0.05
6.93	36.61	0.02	0.02
6.56	28.81	0.05	0.05
5.44	27.41	0.03	0.03
4.81	243.56	0.00	0.00
4.69	113.60	0.01	0.01
-	-	-	-
4.34	82.39	0.01	0.01
-	-	-	-
3.90	52.37	0.01	0.01
3.08	19.97	0.01	0.01

Off-axis

d-Spacing (Å)	Intensity	+d (Å)	-d (Å)
16.60	28.82	0.04	0.04
-	-	-	-

Radially averaged

d-Spacing (Å)	Intensity	+d (Å)	-d (Å)
4.46	55.91	0.02	0.02
3.70	35.53	0.02	0.02

Mat – Parallel

Equatorials

d-Spacings (Å)	Intensity	+d (Å)	-d (Å)
28.08	229.35	0.23	0.23
20.70	36.56	0.71	0.67
13.50	*	*	*

Merdionals

d-Spacing (Å)	Intensity	+d (Å)	-d (Å)
9.86	246.62	0.13	0.13
8.43	80.22	0.03	0.03
6.87	55.29	0.06	0.06
6.47	59.99	0.04	0.04
-	-	-	-
4.77	192.76	0.02	0.02
4.66	*	*	*
4.41	79.81	0.02	0.02
4.32	*	*	*
4.12	67.43	0.03	0.03
3.87	82.71	0.02	0.02
-	-	-	-

Off-axis

d-Spacing (Å)	Intensity	+d (Å)	-d (Å)
16.90	56.74	0.16	0.16
-	-	-	-

Radially averaged

d-Spacing (Å)	Intensity	+d (Å)	-d (Å)
-	-	-	-
-	-	-	-

Disc – Parallel

Equatorials

d-Spacings (Å)	Intensity	+d (Å)	-d (Å)
28.78	102.83	1.48	1.65
19.49	34.70	0.87	0.80
13.39	45.75	0.14	0.14

Merdionals

d-Spacing (Å)	Intensity	+d (Å)	-d (Å)
9.86	231.97	0.01	0.01
8.41	108.54	0.04	0.05
6.87	87.50	0.01	0.01
-	-	-	-
5.47	*	*	*
4.77	227.48	0.00	0.00
4.66	179.09	0.00	0.00
4.41	*	*	*
4.31	152.60	0.00	0.00
-	-	-	-
3.89	134.73	0.00	0.00
-	-	-	-

Off-axis

d-Spacing (Å)	Intensity	+d (Å)	-d (Å)
16.94	64.60	0.03	0.03
6.44	86.22	0.05	0.05

Radially averaged

d-Spacing (Å)	Intensity	+d (Å)	-d (Å)
-	-	-	-
-	-	-	-

Table 6.1: The measured reflections from the fibre diffraction shown in Figure 6.4.

The 4.81 Å meridional reflection is consistent with a hydrogen bonding separation of ß-strands and the ~28 Å equatorial, too large to arise from a spacing between adjacent ß-sheets likely occurs due to the length of the αSß1 peptide length of 28 Å (8 residues x 3.5 Å). Always found on the equator this reflection must occur from a repetitive spacing perpendicular to the fibre axis and so considering a nanotube assembly the ß-strands can occur in or out of plane from the tube wall. Where αSß1 is found to be a peptide amphiphile, as shown in Figure 6.5, it is likely to adopt an out-of-plane configuration in the nanotube wall as also concluded elsewhere (Valery et al., 2011). This is consistent with the fibre diffraction data and is further consistent with the amphiphilic nature of the molecule, whereby two αSß1 peptides may stack end-on-end within the wall to make a bilayer of a thickness of ~56 Å.

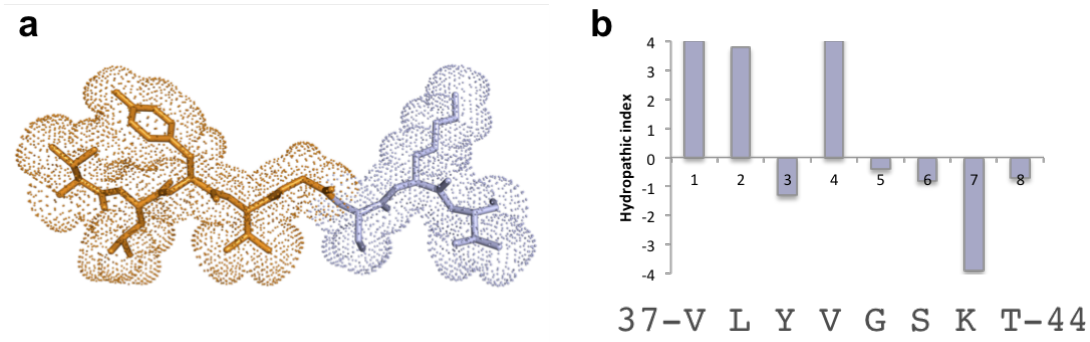


Figure 6.5: The amphiphilic nature of the αSynß1 peptide, hydrophobic residues are shown in orange, hydrophilic in blue (a). The hydropathic index is shown (b) according to the Kyte and Dootlittle hydropathy index (Kyte and Doolittle, 1982). Graphics generated in PyMol (DeLano, 2002).

6.2.3 Supramolecular models of assembly

Based on preliminary interpretation of the XRFD and TEM data a nanotubular morphology can be envisaged as shown in Figure 6.6. Both theoretical models are based on the observation of a tape-like amphiphilic structure twisting into a helical tube. The first model describes the 'closed' mature tube (a) and the second a derivation of this as an open tube (b) or as described by Shao *et al.,* a helical tape (Shao et al., 2011). The precise arrangement within the walls of the nanotube would depend on the arrangement of αSß1 molecules within the constituting amphiphilic tape, but the separation between peptides along the nanotube long axis is presumed to be ~4.7 Å.

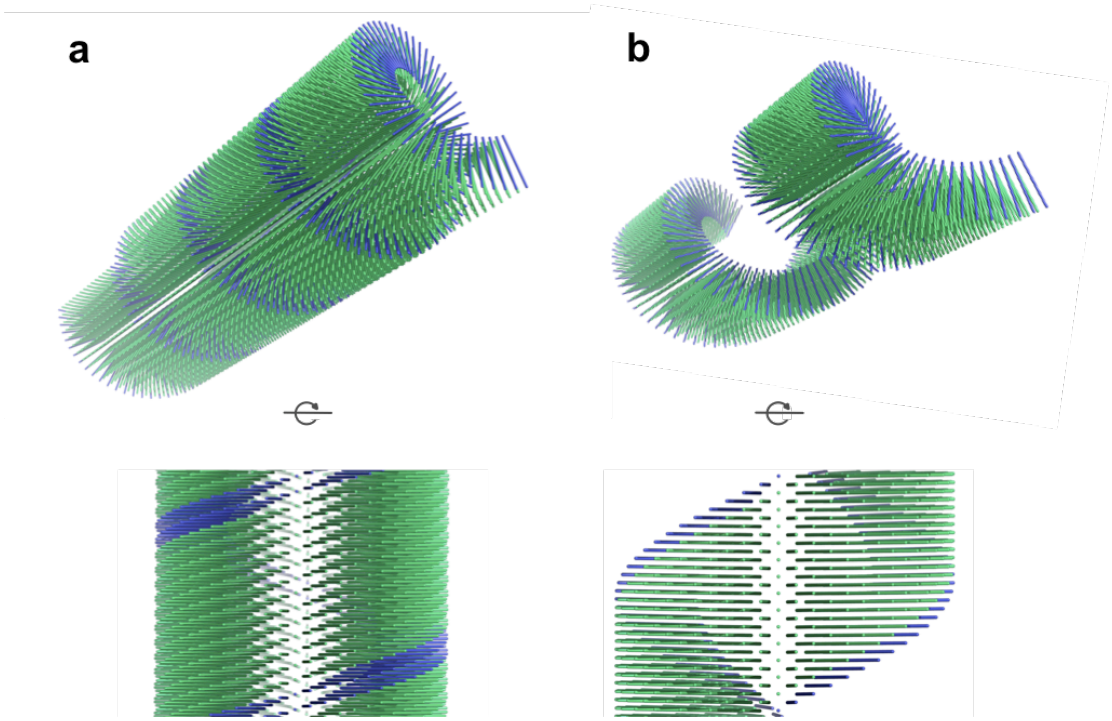


Figure 6.6: A predicted theoretical model of the arrangement of αSß1 peptides to create a nanotubular morphology, as observed by TEM in Figure 6.3. Each line represents a stacked peptide molecule, whereby the separation between peptides along the long axis is presumed to be ~4.7 Å. Peptides are periodically coloured blue for reference. The mature tube (a) is shown as well the helical tape morphology (b). Graphics generated in Pymol (DeLano, 2002), not to scale.

Two possible packing modes for αSß1 are shown in Figure 6.7. If the sample consisted of only flat tape structures then the second architecture (b) would result in the 4.81 and 9.95 Å reflections occurring on the equator and meridian respectively, contrary to what is observed. However extension of this architecture into a nanotubular morphology with a small helical pitch could make the 4.81 and 9.95 Å separations approximately parallel and perpendicular respectively to the nanotubular long axis, consistent with the fibre diffraction. If the sample consisted of only flat tape structures the first architecture (a) would result in a fibre diffraction pattern consistent with the experimental data, the formation of nanotubes from tapes of this architecture could again reproduce the fibre diffraction data but a large helical pitch would be required to make the 4.81 and 9.95 Å separations approximately parallel and perpendicular respectively to the nanotubular long axis.

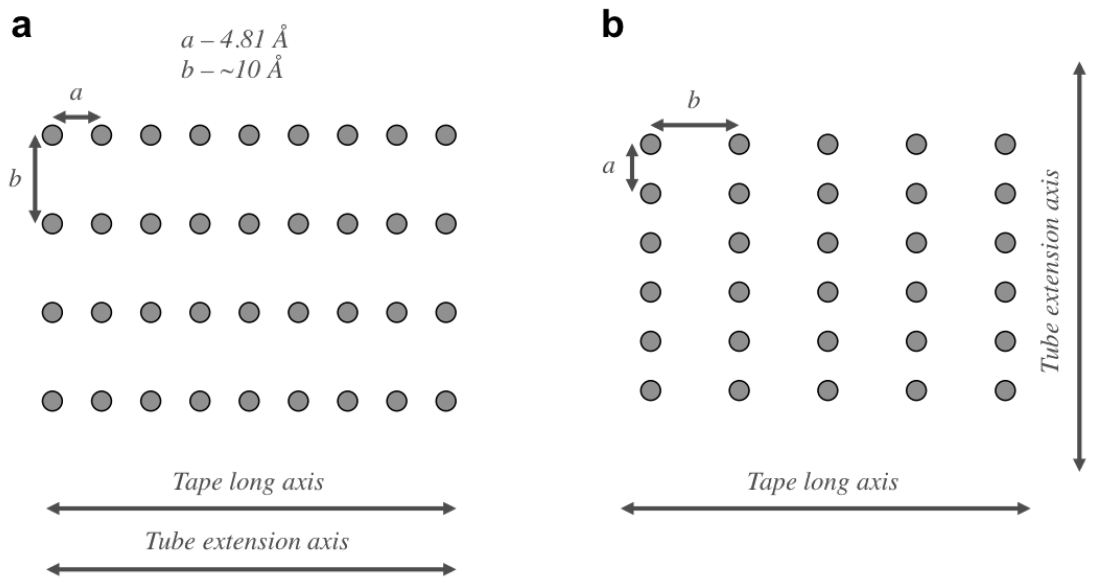


Figure 6.7: Two possible architectures present in the tape structures of αSß1. The long axes of the peptides are perpendicular to the plane to the page and are shown in a parallel 4.81 Å stacking regime (a) or a perpendicular 4.81 Å stacking regime (b).

In both cases, it might be expected that the 4.81 Å peptide separation is tilted away from the long axis of the nanotube. In systems that are reported to have such a supramolecular structure the off axis separation equivalent to the 4.81 Å here is seen to occur as an off axis reflection (Mehta et al., 2008). Whilst this is not the case here, the 4.81 Å reflection must arise from a 4.81 Å separation running parallel to the nanotube long axis. Thus, in either architecture shown in Figure 6.7, the helical pitches must be acutely large close to 90° or 0°. To investigate the exact packing of αSß1 peptides within the nanotube wall, and understand which of the packing architectures discussed is accurate, the fibre diffraction pattern reflections were indexed such that a model of the unit cell constituting the nanotube wall could be constructed.

6.2.4 Molecular models of assembly

The unit cell indexing was performed using the reflections from the film-textured mat (beam parallel) where this was shown to have a fibrillar texture but also resolve the greatest amount of equatorial and meridional information. The putative unit cell indexing of these reflections is shown in Table 6.2. The indexed cell was found to be able to accommodate eight peptide

molecules corresponding to the amphiphilic bilayer of the nanotube wall. The peptide was modelled in an ideal parallel ß-sheet geometry as might be expected for a cross-ß system. The ß-sheet arrangement was assumed to be parallel to maintain the amphiphilicity of the nanotube wall. For simplicity, half of the bilayer was modelled where *c* = 30.00 Å as shown in Figure 6.8a. The cell dimensions *a*9.62, *b*19.88, *c*30.00 Å correlate with the separation of peptide molecules (2 × 4.81 Å) along the fibre axis, ß-sheet separation (2 × 9.94 Å) and the length of the peptide molecule (30.00 Å).

Asymmetry is present along the *a* dimension by a 180 rotation about the *c* axis. The model is able to rationalise the anisotropic growth of the tape-like structures and formation into a helical nanotube where the structure shown in Figure 6.8 is stabilised by three discrete sets of interactions aligned to the three crystallographic dimensions. ß-strands are stabilised along the *a* dimension by interamide hydrogen bonding (b) and perpendicular to this along the *b* dimension adjacent ß-sheets are stabilised by domain swapped Lys-Ser hydrogen bonds and Tyr aromatic interactions (c). The amphiphilic bilayer along the *c* dimension is stabilised by hydrophobic-hydrophobic interactions (d).

αSß1 unit cell

Dimensions	**a (Å)**	**b (Å)**	**c (Å)**	**α = ß = γ**
	9.62	19.88	60	90
Observed / Å	**h**	**k**	**l**	**Calculated / Å**
28.08	0	0	2	30.00
20.70	0	1	0	19.88
13.50	0	1	3	14.10
16.90	0	1	2	16.57
9.86	1	0	0	9.62
8.43	1	1	0	8.66
6.87	1	2	0	6.91
6.47	1	0	7	6.40
4.77	2	0	0	4.81
4.66	2	1	1	4.66
4.41	2	1	4	4.46
4.32	2	2	0	4.33
4.12	2	0	7	4.19
3.87	2	3	0	3.89

Table 6.2: Putative unit cell indexing of the reflections collated from the various textures of αSß1.

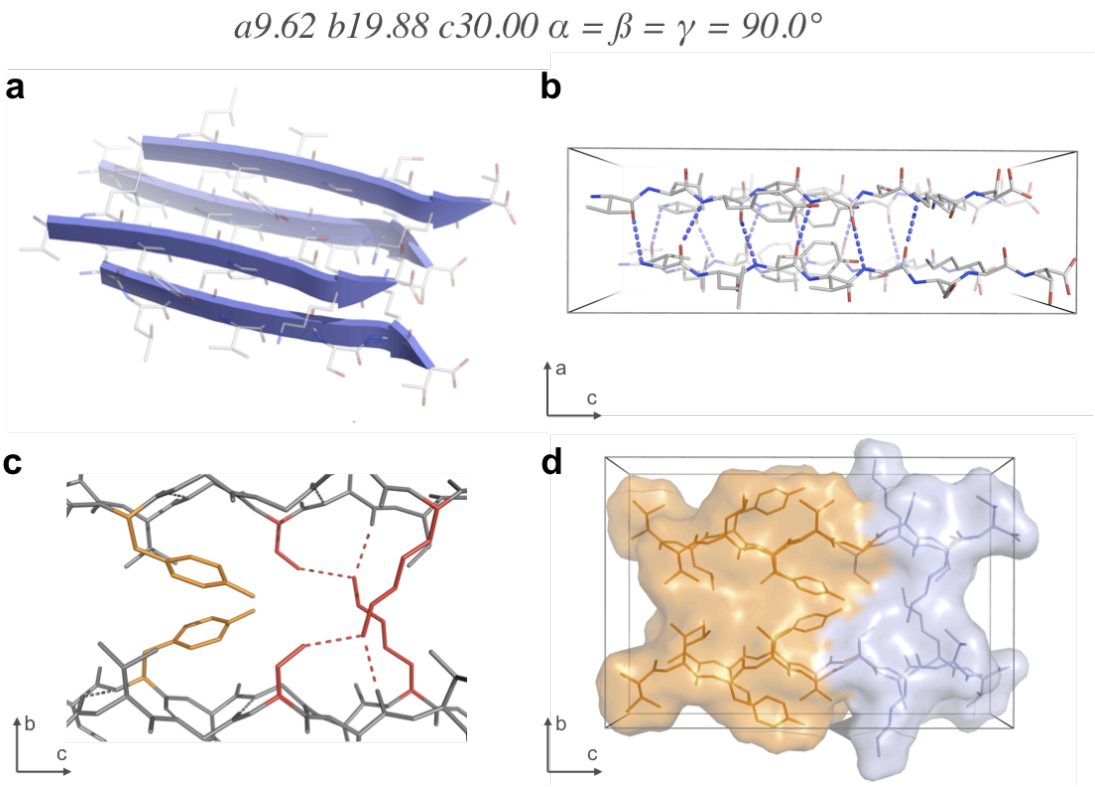


Figure 6.8: A model of the αSβ1 peptide packed into the predicted unit cell (a). Interamide hydrogen bonding (blue lines) stabilises β-strands along the fibre axis a dimension (b), domain swapped inter lysine-serine hydrogen bonding (red) and aromatic interactions (orange) stabilising the b dimension (c) and an amphiphilic polarity stabilising the c dimension (d). Graphics generated in PyMol (DeLano, 2002).

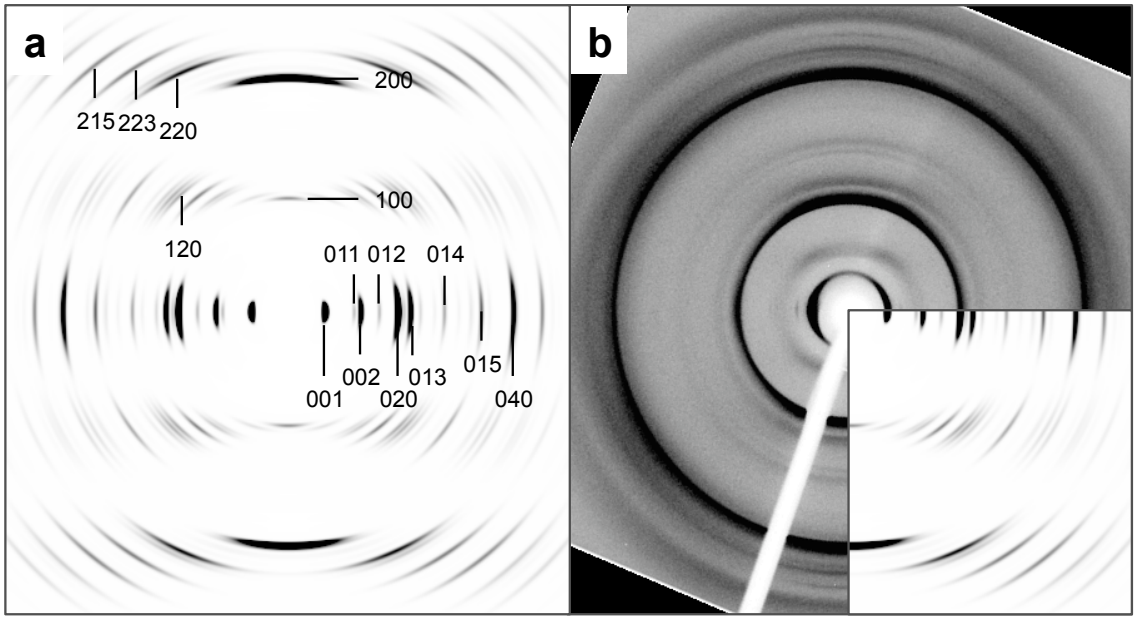
Where hydrogen bonding should be the most directional and strongest of the interactions present this will promote one-dimensional growth along the *a* dimension and so likely corresponds to the long-axis of the tapes - consistent with the tape architecture shown in Figure 6.7a. Then the domain swapped Lys-Ser and aromatic interaction will stabilise the broadening of the tape. Finally, the amphiphilic nature of the unit cell will encourage a stable bilayer to form, with no need for additional stabilisation, thus defining the thickness of the tape. As a result the edge of the tape will present the *bc* plane (see Figure 6.8a and c) to the solvent which could finally be minimised by nanotube formation.

6.2.5 X-ray fibre diffraction simulation structure validation

The molecular model presented in Figure 6.8 is consistent with the fibre diffraction data in as much as the unit cell dimensions and the internal atomic separations but for full validation simulated fibre diffraction was performed as shown in Figure 6.9.

The major reflections of the experimental fibre diffraction pattern are present on the simulated pattern, although the relative intensity and axis alignment of many of the reflections are disparate from the experimental reflections. The equatorial reflection of 20.7 Å and meridional reflection of 4.66 Å are not reproduced in the simulated pattern, although where they are found to index to the unit cell (see Table 6.2) it cannot be ruled out that they are absent due to low relative intensity. Indeed, the meridional 6.47, 4.41 and 4.12 Å reflections are present as off meridional and their relative intensity is very low.

Taken together, the simulation indicates that the unit cell and structure is related to the structure in the nanotube wall but discrepancy is apparent. The relative intensities of reflections will be modulated by the structure architecture and side chain conformations within the unit cell and so a search of the structural space within this cell should be able to correct these differences. Finally, where some reflections occur off-axis it is possible that this represents some structural spacings are aligned to the incorrect axis in the model structure. This is consistent with the modelled unit cell arranging into a supramolecular helical structure where the simulation considers the cell in a simple orthorhombic lattice.



	a	*b*	*c*
Unit cell dimension / Å	*9.62*	*19.88*	*30*

Simulated		Miller indexing			Experimental	
Signal / Å	**Axis**	**h**	**k**	**l**	**Signal**	**Axis**
9.62	M	1	0	0	9.86	M
4.81	M	2	0	0	4.77	M
6.91	OM	1	2	0	6.87	M
4.33	OM	2	2	0	4.32	M
9.94	E	0	2	0	n.o.	-
4.97	E	0	4	0	n.o.	-
30	E	0	0	1	28.08	E
15	E	0	0	2	13.5	E
16.57	E	0	1	1	16.9	OA
11.97	E	0	1	2	n.o	-
8.93	E	0	1	3	8.43	M
7.02	E	0	1	4	n.o	-
5.74	E	0	1	5	n.o	-
3.97	OA	2	2	3	3.87	M
3.69	OA	2	1	5	n.o	-

Figure 6.9: Simulated fibre diffraction from the proposed model of αSβ1. Not observed - n.o. Fibre axes are vertical.

6.3 Summary

Nanotubes have been observed for a number of peptide derived self-assembling monomers including Aß(16-22) (Lu et al., 2003, Mehta et al., 2008), *cyclo*[(-D-Ala-L-Glu-D-Ala-L-Gln-)$_2$] (Ghadiri et al., 1993), Lanreotide (Valery et al., 2003), A_6K (Bucak et al., 2009) and NDI-lysine amphiphiles (Shao et al., 2011). However, of these systems the morphology and scale reported here is unique considering the simple linear peptide sequence that is used.

The αSß1 peptide assembles into flat tape-like structures with a repetitive separation of 4.77 Å along the tape long axis. These tapes assemble over huge length scales of 300 – 400 Å in width and indeterminate length. They are stabilised by hydrogen bonding along their long axes and peptide packing stabilises the lateral tape width, whilst the amphiphilic nature of the peptide causes the thin bilayer tape structure. To further stabilise the structure these tapes may arrange helically to form nanotubes, observed to have complex morphologies where the tape-like structure can be separated along its long axis. The assembled nanotubes exhibit ordered fibre diffraction of a quality only comparable to a few other systems indicating the high degree of order in these systems. This indicates that the nanotube wall consists of a single bilayer tape-like structure, which can be modelled on the cross-ß structure of amyloid proteins. The arrangement of this architecture within the nanotube wall creates a novel cross-ß architecture. At the time of writing this analysis represents the most robust analysis of a nanotubular peptidic supramolecular assembly. A similar supramolecular architecture has only ever been proposed by Mehta *et al.* for Aß(16-22) (Mehta et al., 2008) although in a course representative model. Though the novel cross-ß architecture cannot be validated within the unique texture of the nanotube, future work will address this, either by development of methods to simulate such a complex long-range texture or by the use of other structural techniques including solid state nuclear magnetic resonance spectroscopy.

This work represents the beginnings of attempts to truly link the molecular structure of a self-assembled peptide molecule to its supramolecular assembly. A few models for peptidic supramolecular assemblies have been proposed (Mehta et al., 2008) but are not validated by comparison of simulated structural data to the experimental structural data. It is demonstrated here that it is possible to find a molecular architecture that is related to the supramolecular assembly but great care must be taken in doing so. It is an enticing challenge to truly understand the link between this sort of molecular model and the supramolecular assembly it forms. The methodology here represents a step in the direction for using fibre diffraction and TEM to do so, but also an exquisite example of the elaborate and yet elegant architecture that forms to create such a supramolecular assembly.

7 Results and Discussion: The self-assembly mechanism and fibrillar structure of a bromonaphthalene conjugated dipeptide

7.1 Introduction

7.1.1 Current understanding of LMWG self-assembly

One class of low-molecular weight gelator (LMWG) molecule that have the self-assembly potential to form fibrils capable of gelling a solvent, are the naphthalene conjugated dipeptides, reported previously by Yang *et al.* (Xu et al., 2007) (see Chapter 2). A substantial amount of characterisation of these and related systems has been conducted. One such large body of research surrounds investigations into Fluorenylmethyloxycarbonyl chloride (Fmoc) based LMWGs (see Chapter 2), in particular Fmoc-Phe-Phe (Smith et al., 2008). This Fmoc based LMWG is able to gel water and was reported to have a dramatically shifted pKa of its carboxyl terminus to 9.9. It is well established that the self-assembly of this system occurs when the pH of the environment is below the pKa of the carboxyl terminus, thus explaining its ability to gel solutions below a pH of 9.9 (Tang et al., 2009). This phenomenon has been experimentally verified in a number of other aromatic conjugated LMWGs (Adams et al., 2010b, Chen et al., 2010c), but questions remain over the driving forces of assembly and also the structure of the assemblies formed.

7.1.2 The driving forces of self-assembly

Highlighted in Section 2.4, the intermolecular interactions stabilising non-covalently bonded systems range from electrostatic interactions and hydrophobic effects to hydrogen bonding. Where naphthalene and Fmoc conjugated dipeptide LMWG have a high aromatic content, they are also able to develop π-π stacking interactions. Although hydrogen bonding is also

acknowledged as an important stabilising interaction in these systems, emphasis is most widely placed on the importance of π-π interactions between aromatic groups driving assembly (Jayawarna et al., 2006, Toledano et al., 2006, Yang and Xu, 2004, Zhou et al., 2009). A view where side chain hydrogen bonding and especially peptide backbone hydrogen bonding with co-operative π-π stacking is probably more accurate (Mahler et al., 2006, Mart et al., 2006, Xu et al., 2007). Indeed, some studies on Fmoc-Leu-Gly indicate that the initial assembly is driven by aromatic interactions which once established are unchanged, whilst other structural changes continue to occur as indicated by the continued evolution of circular dichroism (CD) signals (Adams et al., 2009).

It is well recognised that for a supramolecular assembly to form fibrils which are capable of gelling solvent, the non-covalent interactions stabilising the assembly must be one-dimensional (Dastidar, 2008). Aromatic interactions may be attributable to driving the initial stages of self-assembly, but in the final structures are they directional enough to stabilise the one-dimensional bonding networks that promote anisotropic growth? The observation by Gazit *et al*. of nanotubes formed by di-phenylalanine is an example that illustrates π-π interactions can be exceedingly directional however, though these stabilising forces lead to anisotropic supramolecular assemblies (Reches and Gazit, 2006), the aromatic-aromatic interaction is still the subject of study where it is not fully understood. Studies of benzene-dimers within the Cambridge Structural Database (CSD) and Protein Data Bank (PDB) reveal three common π-π geometries: parallel-displaced (PD), T-shaped edge-to-face (TEF) and eclipsed face-to-face (EFF) as shown in Figure 7.1. The PD and TEF conformations are preferentially adopted, perhaps suggesting that from this example, that although there is a directionality in this interaction atomistically these configuration are not aligned in the same repetitive manner as the hydrogen bonding in polypeptides.

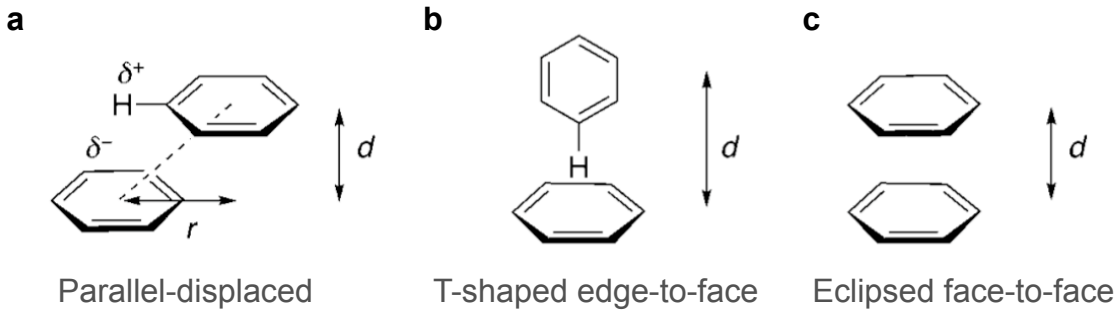


Figure 7.1: Three observed π-π interaction geometries for benzene-benzene dimers. Figure reproduced with permission from (Salonen et al., 2011).

Relevant to the relation of these forces to hydrogelation, the di-phenylalanine nanotubes do not gel the solvent they are dissolved in (Reches and Gazit, 2006), and there are many examples for other LMWGs reported that also fail to form gels (Chen et al., 2010c, Muro-Small et al., 2011, Xu et al., 2007). Interestingly, in these cases, it is not reported whether the non-gelling systems are still self-assembled and so it can be questioned as to whether the stabilising interactions that drive self-assembly are distinct from the interactions that govern gelation.

7.1.3 Current understanding of the structure of LMWG

The current structural insight into peptidic LMWGs is discussed in detail in Chapter 2. These generally describe antiparallel stacks of dipeptide chains in a ß-strand conformation stabilised by interamide hydrogen bonding. π-π stacking interactions are described and can be seen to extend in the same direction as the interamide bonding (see Figure 2.10 and 2.11). A caveat is that, although these models are based on robust structural and biophysical constraints, the final structures presented are ultimately not validated by comparison of simulated to the experimental structural data from which they were generated. Further, the models described draw conclusions about interatomic distances without directional information i.e. Data is collected from unaligned samples of gel in wide-angle X-ray scattering (WAXS) measurements. With a thorough structural and biophysical characterisation of 2-(2-(2-(6-Bromonaphthalen-2-yloxy)acetamido)propanamido)-3-methylbutanoic acid (herein referred to as Br-AV) and simulation of structural information from potential models this work aims to produce a

representative and experimentally validated model of an aromatic conjugated dipeptide in the fibrillar phase.

7.1.4 Current understanding of the viscoelastic properties of hydrogels

The viscoelastic properties of hydrogels are widely appreciated (Adams et al., 2009). However, a comprehensive rheological characterisation of these properties is often neglected, and viscoelastic properties are confined to a simple description of gelation or no gelation based on 'table-top rheology'. No firm link has been described between molecular structure and gelation propensity. Further, where detailed rheological characterisation is reported only in a few examples (Chen et al., 2010a, Chen et al., 2010c) of LMWG studies, the subtleties between molecular structure and hydrogel viscoelastic properties has not been explored. It has been noted that there are particular geometrical constraints on the dipeptide molecule that dictate gelation capacity (Xu et al., 2007). However, these constraints most likely dictate the supramolecular assembly and organisation of the LMWG, which then affects self-assembly, which in turn affects gelation potential. Where more involved characterisations of viscoelastic properties of peptidic LMWGs have been made, currently, no tangible links have been made between viscoelastic properties and structure.

Clearer identification of the driving forces of assembly, how these interactions are able to stabilise the growth and final structure of the fibrillar phase and ultimately information regarding the architecture of structural elements within the fibres it may be possible to correlate structural properties to the viscoelastic properties of these gelators, thus bringing rational LMWG molecule design closer to reality.

7.1.5 Experimental aims

The assembly mechanism of peptidic LMWGs that assemble via a protonation mechanism is widely accepted and understood based on theoretical and final gel properties (Tang et al., 2009).

The structure of LMWGs in the final fibrillar state has been probed previously but the structures determined can only be considered as models (Hughes et al., 2011, Smith et al., 2008, Xu et al., 2010) and no successful attempt has been made to truly correlate these models with macromolecular and viscoelastic properties. The structural details determined here for Br-AV may be applicable to other LMWG systems, certainly its characterisation serves as a good methodological and structural model for better understanding the assembly and final structure of this class of LMWG.

With the development of a method for lowering the pH of the solvent, the dipeptide is dissolved in a homogenous and kinetically controllable manner (Adams et al., 2009); for the first time the assembly of a LMWG can be monitored in real time. This presents the opportunity to monitor time dependent structural changes and gain insight into what forces drive the self-assembly of this system. Additionally, through precise control of the final conditions in which the fibrillar assemblies stabilise their structure may be probed revealing details about the structural architectures present at various mechanistic points of the self-assembly and gelation process. In this manner, the hydrogelation process may be understood in detail as well as an attempt made to correlate structure and stabilising interactions to the mechanical properties of the gel formed.

This chapter reports on the findings published by Chen *et al.* on the detailed characterisation of the assembly and structure of the low-molecular weight gelator Br-AV (Chen et al., 2010a) and expands on the analysis of the data collected by K. L. Morris.

Dr. D. J. Adams. (University of Liverpool) is gratefully acknowledged for synthesising and providing Br-AV. Dr. L. Chen (University of Liverpool) is acknowledged for the collection and supply of pH decrease measurements.

7.2 Results and Discussion

7.2.1 Homogenous and reproducible lowering of pH using Glucono- -lactone

Naphthalene conjugated dipeptides have previously been reported to form gels at acidic pH (Xu et al., 2007). Using glucono-δ-lactone (GdL) to homogenously lower the pH of the solution containing Br-AV (Adams et al., 2009) the pH was lowered to 3.1 as shown in Figure 7.2a. The pH drop arising from the production of *D*-Gluconate is at first rapid and subsequently steady. A concentration of 14.42 $mgmL^{-1}$ GdL consistently lowered the pH to 3.1 and with reproducible kinetics (Adams et al., 2009) (D. J. Adams, communication).

The apparent pKa of Br-AV is 5.8 (Chen et al., 2010c). At pH values above this the carboxy-terminus will be in the carboxylate form. At these pH value solutions have been observed to form micelle-like structures (Chen et al., 2011), presumably stabilised by ionic interactions of these groups with water. A pH of 5.8 is reached by the solution between 30 – 40 seconds. By this time the carboxylate terminus of the free dipeptide in solution will be protonated to a carboxyl termini abolishing the negative charge of this group as shown in Figure 7.2b.

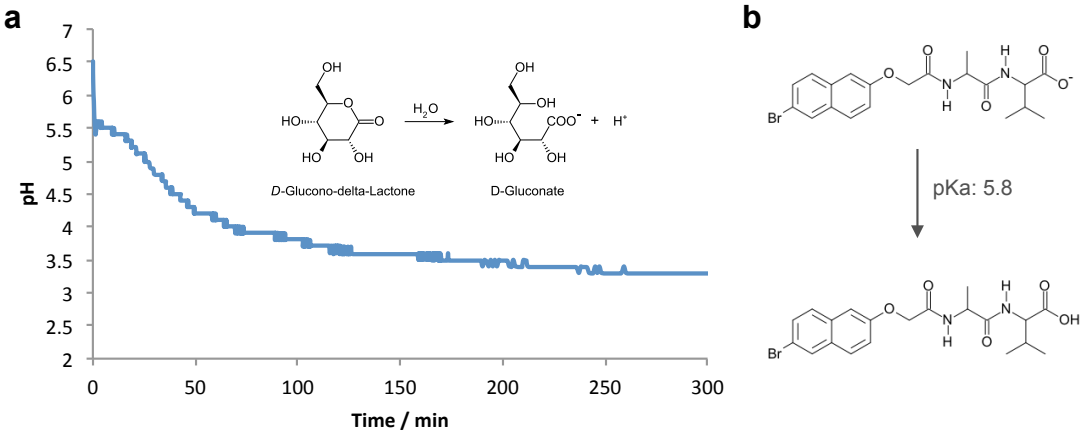


Figure 7.2: The lowering of the pH of a solution of Br-AV (5 $mgmL^{-1}$) at an initial pH of 10.7 using 14.42 $mgmL^{-1}$ of GdL (a) and the associated theoretical ionisation state of Br-AV based on this (b). pH data collected and provided by Dr. Lin Chen (University of Liverpool).

7.2.2 Sampling from hydrogels

Assuming reproducible handling of the GdL during the initial rapid pH decrease then the subsequent time course of acidification and hydrogelation may be followed. Initially though, solutions were left to gel overnight via this pH drop and later were sampled to investigate their final gel properties. However, sampling from preformed hydrogels was found to not effectively remove the fibrillar material from overnight formed hydrogels. After gel penetration by a pipette tip, a negative impression is left in the gel into which fluid flows. Sampling this material results in transmission electron microscopy (TEM) grids that have a high ratio of salts and precipitate (presumed to be NaOH and GdL) to peptidic material as shown in Figure 7.3a. Analysis of this material by CD results in spectra that resemble GdL only. Figure 7.3b shows the spectral signature for the material sampled from a preformed gel is comparable to the spectral signature of GdL alone (Jia et al., 2009). Furthermore, no absorbance spectra can be derived from the high tension voltage (HT[V]) as would be expected for the naphthalene chromophore (data not shown). It was investigated to see whether hydrogels formed *in situ* would better enable the investigation of the material formed in the gel matrices. The following describe experiments conducted using *in situ* methodologies whereby measurements are made directly on gels and not on samples removed from preformed gels.

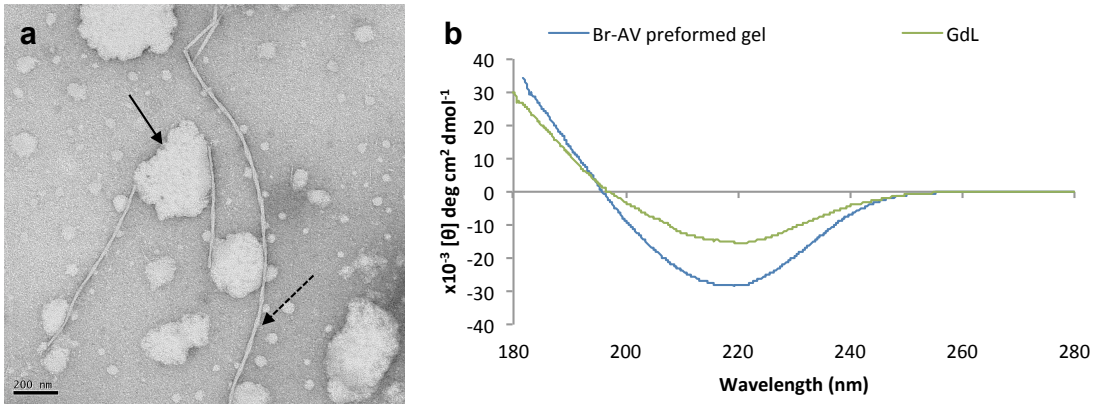


Figure 7.3: Sampling problems encountered when using pre-formed hydrogels. Sampling from preformed gels for TEM results in low coverage of fibrillar material (dashed arrow) and high degree of GdL precipitate (solid arrow) (a). CD analysis of this sampling yields a GdL CD spectrum (b).

7.2.3 Monitoring the self-assembly of Br-AV by TEM

After addition of GdL the macroscopic and morphological structure of the gel matrix was probed at specific time points as shown in Figure 7.4. At the point of GdL addition any material observed is unstructured but by 40 mins (pH 4.2) this has adopted a fibrillar morphology. These fibrils observed at low coverage have a minimum width of 12.9 nm (SD ±1.0 n=10). At 80 mins (pH 3.8) the incidence of these fibres increases with some appearing wider with striations parallel to the fibre axis, clear evidence of lateral association. At this time point a laterally associated thick fibril was found to constitute thin filaments with a width of 12.0 nm (SD ±0.5 n=10). Instances of highly laterally associated fibrils are observed at all time points up to complete gelation. At 120 mins (pH 3.7) some fibrils are unstained, presumably because they are so large they cannot be negatively stained. As gelation progresses the fibril coverage and overlap incidence increases, high order lateral association appears to plateau as fibril overlapping increases. It is noteworthy that no helical structures were observed as was previously reported for naphthalene conjugated dipeptides although these studies were of LMWG's with the sequences, Gly-Gly, Gly-Ala, Gly-Ser (Xu et al., 2007). This would suggest that helicity and chirality are governed by the peptide portion of LMWG's.

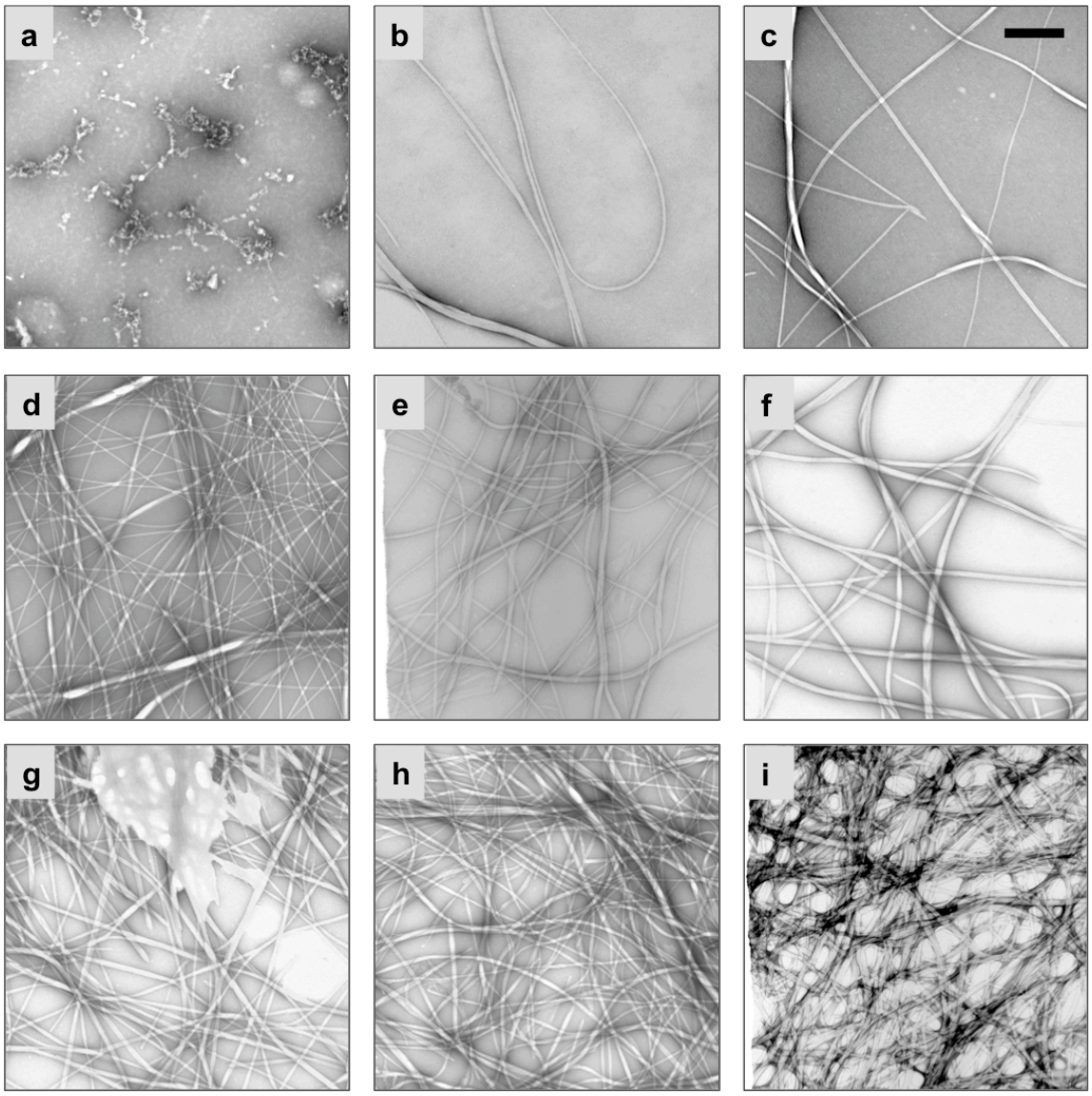


Figure 7.4: The development of fibrillar structures and gel matrix formed over time by the dipeptide Br-AV following the addition of GdL (14.42 mgmL^{-1}). Micrographs were taken at 0 (a), 40 (b), 80 (c), 120 (d), 160 (e), 200 (f), 240 (g), 280 (h) and 400 mins (i). Scale bar represents 400 nm. Figure reproduced from (Chen et al., 2010a).

7.2.4 Monitoring the self-assembly of Br-AV by naphthyl fluorescence

The fluorescent emission (λ_{ex} – 265 nm) of the naphthyl group was monitored during the course of assembly and gelation as shown in Figure 7.5. The fluorescent emission from the naphthalene group was found to increase with self-assembly, consistent with a burial of the naphthalene groups allowing for increased fluorescence due to decreased quenching from water. These observations are similar to the phenomenon of increased aromatic fluorescence upon protein folding due to burial and desolvation tyrosine and tryptophan residues (Weinberg, 1988). It is

interesting that Fmoc dipeptides exhibit Stokes shifting in their fluorescent emission spectra upon assembly and this is attributed to aromatic stacking (Hughes et al., 2011, Jayawarna et al., 2006, Smith et al., 2008), however this is not observed here. These data might thus be giving indication of a lack of aromatic stacking, however in light of evidence from CD this is clearly not the case and close aromatic interactions are present, as discussed in the following sections.

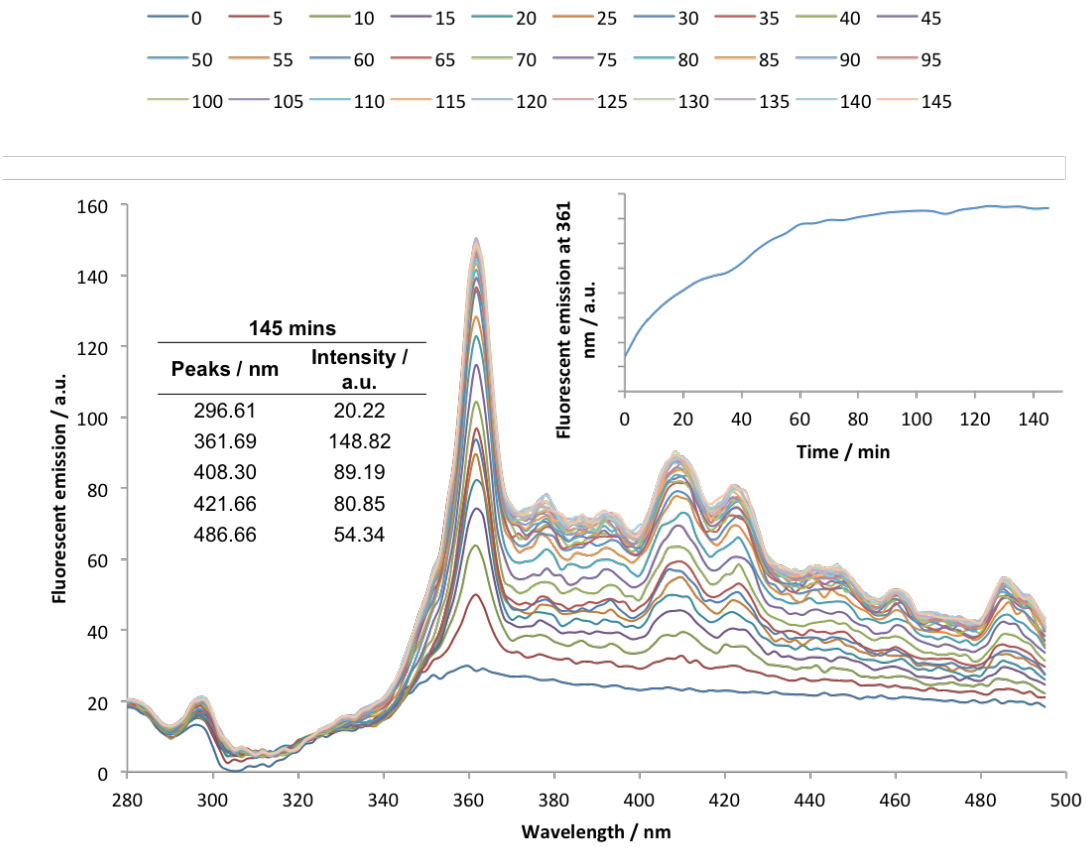


Figure 7.5: The increasing fluorescent emission spectrum of assembling Br-AV (5mgmL^{-1}) at an initial pH of 10.7 after the addition of 14.42 mgmL^{-1} of GdL at 0 (dark blue), 50 (light blue) and 145 (orange) mins. λ_{ex} – 265 nm.

7.2.5 Monitoring the self-assembly of Br-AV by CD

To further investigate the role of assembling interactions and the development of structure, CD was used to probe the system with self-assembly and gelation. As shown in Figure 7.3 CD spectra collected from preformed gels are only able to sample the GdL left in solution after gelation and thus to collect CD spectra from assembling dipeptides gelations were monitored *in situ*.

GdL background

Since the background CD signal due to GdL could potentially contribute to the spectra collected of the assembling Br-AV, this was first investigated. Figure 7.6 shows the CD spectra of hydrolysing GdL over time. The transition at 222 nm blue shifts by 4.8 nm and decreases intensity by 67 %. The CD signals arising from GdL between 210 – 240 nm are about two orders of magnitude less than the CD signals arising from the dipeptide and so despite their changing CD profile are thus not accounted for in subsequent CD analysis.

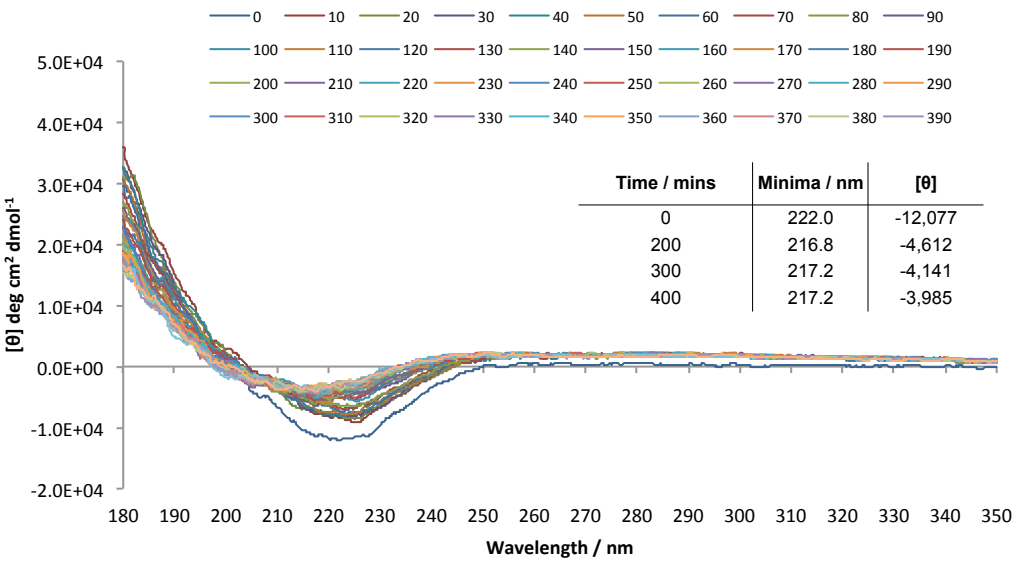
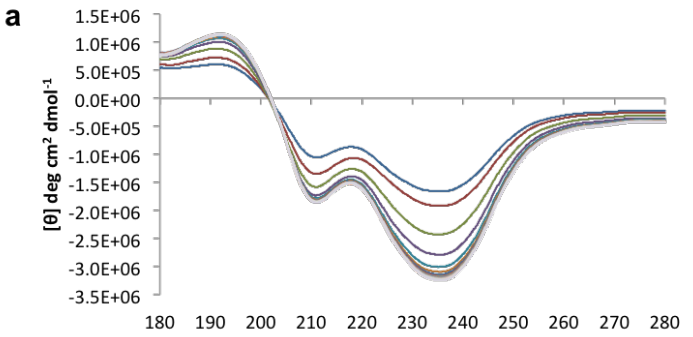


Figure 7.6: The background from GdL at 14.42 mgmL^{-1} in solution at a starting pH of 10.7 is dynamic but the intensity negligible compared to the signals of the dipeptide molecule.

In situ monitoring of the assembly of Br-AV by CD

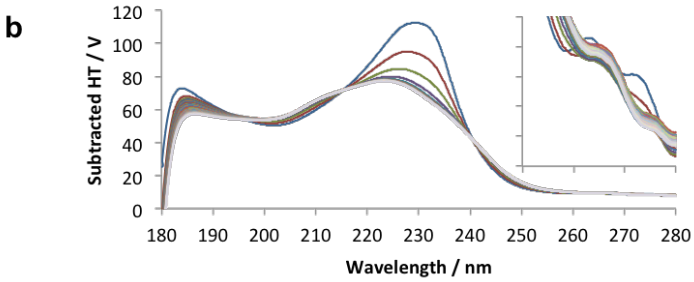
Figure 7.7 shows the spectral characterisation of assembling Br-AV after the addition of GdL. The CD signature of the dipeptide is present immediately at 0 mins (blue) and the peaks increase over time (Figure 7.7a), consistent with an increase in the structure and chirality they are reporting. The presence of the CD signature of Br-AV at 0 mins suggests that the initial kinetics of assembly were faster than it was possible to begin measurements. This is in agreement with the kinetics of the pH drop induced by this concentration of GdL where the pKa of the dipeptide is reached in 40 seconds; faster than it is possible to begin measurements on the assembling system (See Section 7.2.1). It is interesting though that by TEM at 0 minutes no fibrillar assemblies are observed (see Figure 7.4a).

The subtracted HT[V] of the sample and the buffer may be used to determine the absorbance of the dipeptide (Figure 7.7b). Comparison to the UV/Vis spectrum of the dipeptide stock supports the validity of this method (Figure 7.7c). The CD and the absorbance spectra are clearly extremely complex and cannot be interpreted simply as one would for a protein or peptide; they may however be rationalised on the basis of the absorbing chromophores present in the system. Two types of chromophores are present in Br-AV, the two amide bonds of the dipeptide backbone and the naphthalene group, with absorbance regions expected to be between 190 – 230 nm and 220 – 330 nm respectively (Fasman, 1996, Singh and Thakur, 1981). Although having been reported to be exceptionally complex (Kourouklis et al., 1982), the major electronic transition moments of naphthalene are shown in Table 7.1 (Singh and Thakur, 1981). Some shifts are observed in the electronic transitions when substitutions are made (Dempsey et al., 2010) but these are found to be comparable to the absorbance bands observed for Br-AV. It should noted that the polarisations of these transitions are known to alter upon substitution but these are not reported or experimentally measured here (Singh and Thakur, 1981).



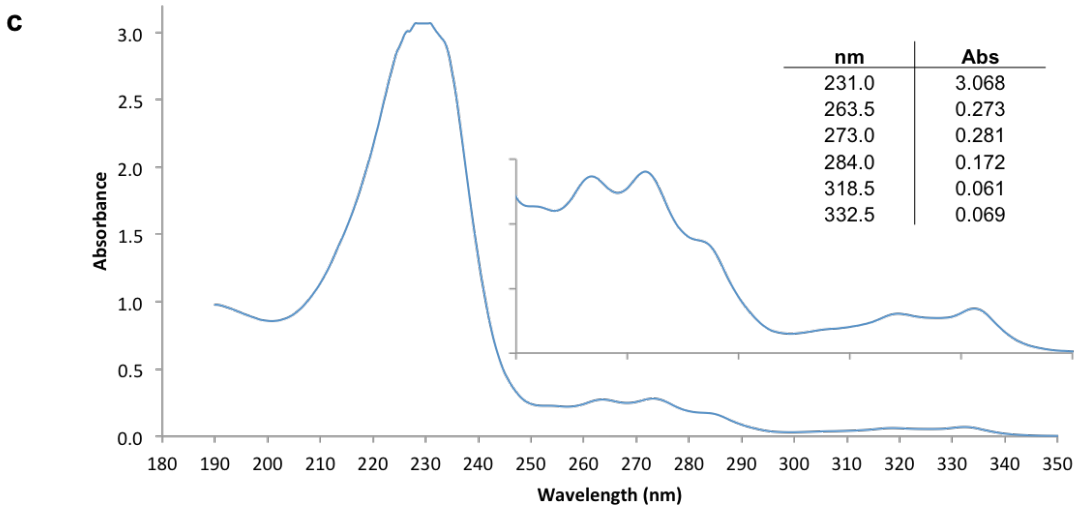
CD

0 mins		600 mins	
nm	[θ]	nm	[θ]
191.7	6.06E+05	191.7	1.14E+06
210.8	-1.05E+06	211.0	-1.86E+06
217.9	-8.64E+05	217.7	-1.51E+06
235.8	-1.65E+06	235.4	-3.24E+06



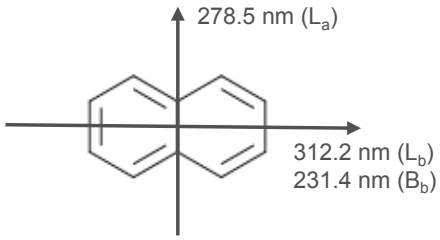
Sub HT

0 mins		600 mins	
nm	Sub HT / V	nm	Sub HT / V
183.9	72.91	186.5	57.08
229.2	112.54	223.3	77.12
262.8	9.65	264.8	9.37
271.5	9.04	*	*



nm	Abs
231.0	3.068
263.5	0.273
273.0	0.281
284.0	0.172
318.5	0.061
332.5	0.069

Figure 7.7: The evolution of CD signals from gelating Br-AV after the addition of GdL. The signature from the dipeptide increases over time with a scan interval of ten minutes (a). The naphthalene chromophore absorbance decreases with time (b). The UV/Vis of the Br-AV stock (25 μgmL^{-1}) is shown for reference, with an inset expanding the near-UV region (c).



Naphthalene Absorbances			6-bromo-2-naphthol Absorbances			Br-AV Absorbances		
nm	Assignment	Angle	nm	Assignment	Angle*	nm	Assignment	Angle**
312	Lb	0	339	Lb	-	332	Lb	-
278	La	90	276	La	-	273	La	-
231	Bb	0	232	Bb	-	231	Bb	-

*Table 7.1: The experimentally determined electronic transition moments of naphthalene (Singh and Thakur, 1981) and the related 6-bromo2-naphthol compound (Dempsey et al., 2010) which may be compared to the electronic transitions of Br-AV. Assignments are made using the Platt notation. *The angle of the electronic transition moments are not available or ** experimentally determined.*

Thus, the 231.0 nm transition observed in the UV/Vis and 229.2 nm in the $HT[V]_{sub}$ of Br-AV can be attributed to the B_b (long-axis) electronic transition moment of naphthalene (Singh and Thakur, 1981). The 235.8 nm transition in the CD spectrum may also be attributed to the naphthalene chromophore (Figure 7.7a). This is an interesting observation where isolated naphthalene is achiral and so does not have an intrinsic CD signal, thus the development of a CD signal corresponding to the B_b naphthalene transition indicates that the naphthalene groups are entering an ordered chiral environment. The ordering of naphthalene groups is likely to be governed by π-π aromatic interactions. This interpretation is corroborated by the concurrent hypsochromic and hypochromic shifting of the absorbance signals determined from the $HT[V]_{sub}$ spectra. This can be likened to the phenomenon of spectral changes observed when denaturing DNA. In natively structured DNA the aromatic bases interact via their π-electron clouds. The UV absorbance of DNA at 260 nm is a consequence of electronic transitions associated with the π-electron orbitals of the bases. In a π-π stacked aromatic structure the number of allowed π-π* transitions is reduced and so the absorbance at 260 nm can be reduced by as much as 40% through a complex series of interactions of the electronic orbitals of aromatic groups (Rhodes, 1961).

Electronic transition moments have been reported for naphthalene in the far-UV region <230 nm (Kourouklis et al., 1982); but, even if present here, are comparably weak as shown from the UV/Vis of Br-AV (Figure 7.7c). Thus, the CD transitions below 220 nm are attributable to the amide bond of the dipeptide backbone and its conformation. A ß-sheet peptide is expected to have positive and negative CD bands at 195 and 217 nm corresponding to the π-π* and n-π* respectively (Townend et al., 1966). The positive 192 and negative 211 nm bands of Br-AV may respectively arise from the π-π* and n-π* transitions associated with ß-sheets. The wavelength of the n-π* transition may be considered quite short for the ß-sheet conformation but has been reported to be highly sensitive to environment and found in other ß-sheet systems to be as short as 210 nm (see Section 3.2.3). Indeed, the evidence for ß-sheet formation is supported by Fourier transform infrared spectroscopy (FTIR) data that shows ß-sheet formation upon gelation with peaks at 1628 and 1679 cm^{-1} (Chen et al., 2010a). The 1679 cm^{-1} peak possibly arises from the presence of an anti-parallel ß-sheet arrangement (Pelton and McLean, 2000). The final pH of the solution can be targeted by the concentration of GdL used, a range of concentrations were investigated using FTIR and it was observed that the signals indicative of antiparallel ß-sheets relatively increased with decreased pH. An additional peak at 1718 cm^{-1} was also ascribed to a protonated carboxyl terminus (Chen et al., 2010a).

It has also been noted that the character of the signals in the CD spectrum are reminiscent of exciton coupling bands (Chen et al., 2010a). Exciton coupling in CD results in the splitting of an absorbance peak into a positive and negative component characterised by a sharp change in signal intensity between the two (Rodger and Norden, 1997). The signs of the exciton couplets are indicative of the chirality of the exciton coupled chromophores chirality (see Chapter 3). The difference in the peak positions of the Bb transition of naphthalene in Br-AV between the CD and UV/Vis may indicate that the 229 nm absorbance band has split into a negative CD component at 236 nm and a low wavelength positive component overlapping with other far-UV transitions (Chen et al., 2010a). The signs of the exciton couplets may represent a left-handed helical arrangement (Chen et al., 2010a, Rodger and Norden, 1997 Oxford university Press).

The low wavelength transitions are also reminiscent of exciton coupling with a negative CD component at 211 nm and a positive component at 192 nm, although from the absorbance spectra it is not clear which absorbance band this arises from.

Linear dichroism (LD) checks were performed to ensure the signals from which these interpretations are made are true CD signals, upon cuvette rotation peak intensities changed by 19 – 22 % but no signal inversion was observed as would be expected for an LD artefact. Additionally, no chiral scattering was observed and the relatively small changes in signal intensity can be attributed to imperfections in the quartz cuvettes or slight misalignment of the cuvette upon mounting.

In situ Br-AV assembly CD analysis

The kinetics of the evolution of each CD peak cannot be analysed simply where the initial stages of self-assembly are not recorded due to the rapid pH decrease to below the Br-AV pKa by 30 – 40 seconds. Despite this, the relative intensities of the development of each signal are shown in Figure 7.8, the 211 nm peak appears to plateau before the almost concurrent plateauing of the 192, 236 and 274 nm peaks; but whether this has structural significant structurally for which parts of the molecule enter into ordered assembly is unclear. If the 211 nm peak arises from the amide bond conformation then it would be expected to concurrently plateau with the 192 nm transition. If the 211 nm transition is the negative couplet of an exciton band it is also not clear why this component would plateau before the positive band. As such, at this stage if the CD bands can be ascribed to independently represent the development of unique structural features these data suggest that each structural element develops simultaneously.

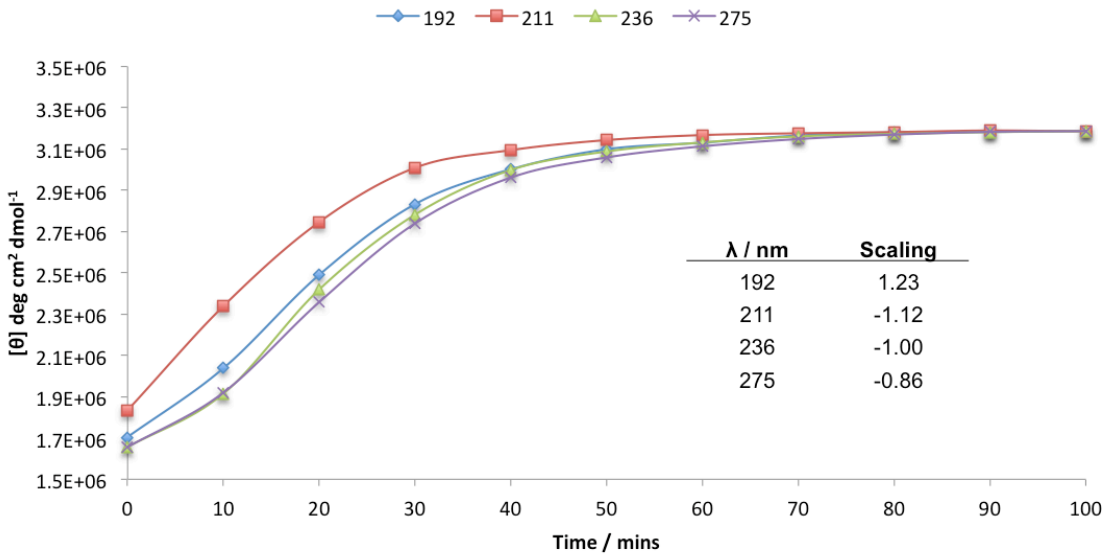


Figure 7.8: The kinetics of CD signal development at 192 (blue diamonds), 211 (red squares), 236 (green triangles) and 275 nm (purple crosses). The data has been scaled for comparison.

Structural geometry interpretation from CD and HT[V]$_{sub}$ measurements

Br-AV exciton coupling

From the HT[V]$_{sub}$ and UV/Vis absorbance spectrum of Br-AV shown in Figure 7.7b-c three transitions can be observed for the naphthalene group. It has already been discussed that substitution affects the polarisation of the transitions associated with naphthalene (see Table 7.1) however for this interpretation the B_b and L_a transitions are assumed to be approximately aligned respectively to the long and short axis of the naphthyl group. The B_b transition is observed at 231.0 nm and the L_a transition at 273.0 nm is split into a second overlapping peak at 263.5 nm. Exciton coupling for chromophores has been demonstrated up to 23 – 50 Å (Matile et al., 1996, Oancea et al., 2003), in particular, naphthalene exciton coupling has been demonstrated in groups spaced up to 6.02 Å (Scholes et al., 1993, Scholes et al., 1998). Thus naphthyl-naphthyl distances cannot be simply described here but each transition can be considered geometrically in its own right to understand the relative geometry and orientation of naphthalene groups in the assembled state.

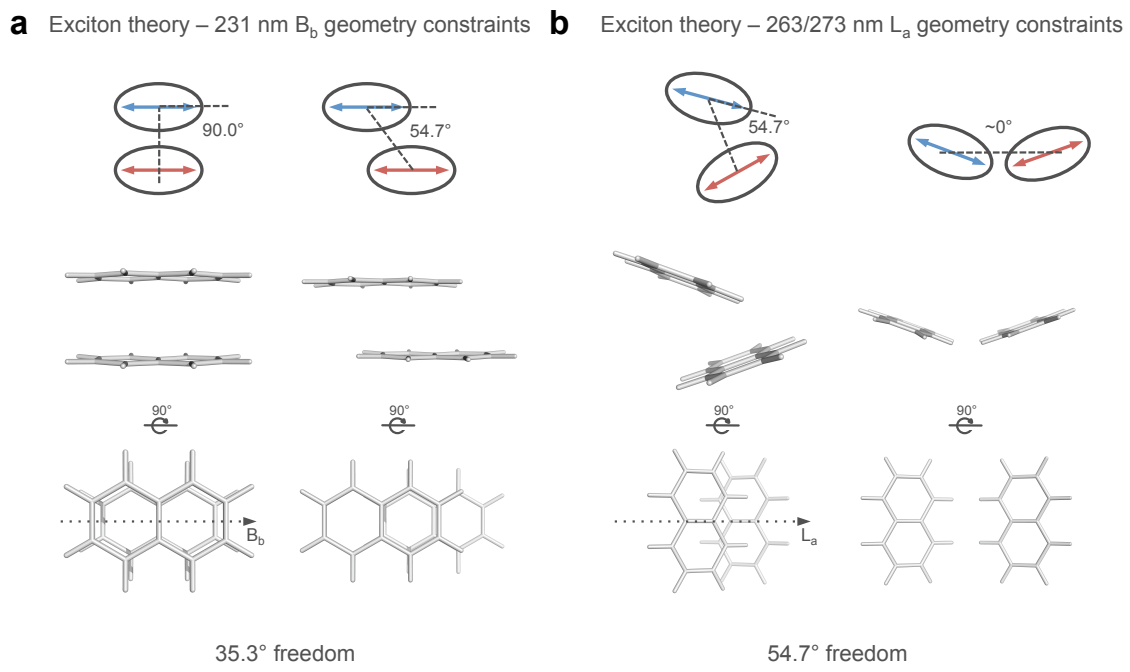


Figure 7.9: The geometrical constraints on the relative arrangements of naphthalene groups within the exciton coupling theoretical framework with respect to the 231 nm B_b (a) and 263/273 nm L_a (b) transitions.

As assembly proceeds the B_b absorption band undergoes a hypsochromic shift of 7.7 nm, according to the exciton coupling model previously described (Chapter 3) consistent with a face-to-face arrangement with regard to the parallel-polarised planar geometry of the B_b long axis transitions. This shift and the lack of band splitting for this transition indicates that the long axes of the naphthalene groups are approximately parallel. The band splitting exhibited by the L_a short axis transition indicates that this transition lies at an oblique angle to its nearest neighbour, but in addition the slight bathrochromic shift of the 263.5 nm L_a component with assembly may indicate an end-to-end geometry for this transition.

The Stokes shifts observed for the assembly of this system upon aromatic interaction are small where shifts may normally be observed on the order of 100 nm for end-to-end (Würthner et al., 2011) and face-to-face (Emerson et al., 1967) arrangements. With regard to determining the geometrical arrangements of the naphthalene groups with respect to the exciton theory this implies that the electron orbital energy splitting is decreased due to the freedom available

through the angles θ and α illustrated in Chapter 3. As such, the relative geometries of naphthyl groups may be determined but only within only the upper and lower geometrical bounds for the B_b and L_a transitions in the face-to-face and end-to-end arrangement. Figure 7.9 shows the B_b transition associated with the long axis of naphthalene must lie parallel to adjacent naphthalene long axes producing a H-aggregate stacking geometry. The hypsochromic shift of this transition indicates that the inclination between the centres of two naphthalene groups with regard to the long axis (θ) must lie between 54.7° and 90.0° (see Chapter 3). The relatively small shift could indicate the groups are inclined close to 54.7° (see Figure 7.9a), but additionally, rotation about α between 0.0° and 90.0° may also contribute to the slight shift (see Chapter 3). The L_a transition associated with the short axis of naphthalene is tilted at an oblique angle relative to adjacent naphthyl short axes to produce the observed band splitting. Additionally the inclination between the centres of two naphthalene groups along this axis (θ) must lie between 0.0° and 54.7° to produce the characteristic red shift of this J-aggregate geometry (see Chapter 3) (see Figure 7.9b).

7.2.6 X-ray fibre diffraction of pH targeted assemblies

To investigate whether the structural architecture changes over the course of assembly pH targeting was used. In doing so, it was possible to probe the structures present at specific time points in the course of the assembly of Br-AV following addition of GdL at 14.42 $mgmL^{-1}$. Table 7.2 shows the range of concentrations of GdL that were used to target the final pH of Br-AV dipeptide solutions. Each GdL concentration successfully induces the hydrogelation of Br-AV where the pH range achieved is 4.5 - 3.1. By allowing the pH of the gels to equilibrate over 24 hours the structures formed will do so in an environment equivalent to specific time points along the time course gelation. Important to these measurements, gel strength as indicated by G' plateau is observed to be at ~50 mins (Chen et al., 2010a).

[GdL] pH relationship				
Time / min	pH	[GdL] / $mgml^{-1}$	Final pH	Equivalalent time point / min
0	6.5	2.94	4.7	34 - 36
40	4.5	4.46	4.0	64 - 74
80	3.9	5.96	3.7	103 - 126
120	3.7	14.42	3.1	>1335
160	3.6			
200	3.5			
240	3.4			
280	3.3			
320	3.3			

Table 7.2: Correlation between the pH and the time after which GdL is dissolved at 14.42 $mgmL^{-1}$ (left). Correlation between the concentration of GdL and the final equilibrated pH of the solution and equivalent time point when using 14.42 $mgmL^{-1}$ GdL in a time course (right). Final pH data taken from (Chen et al., 2010a).

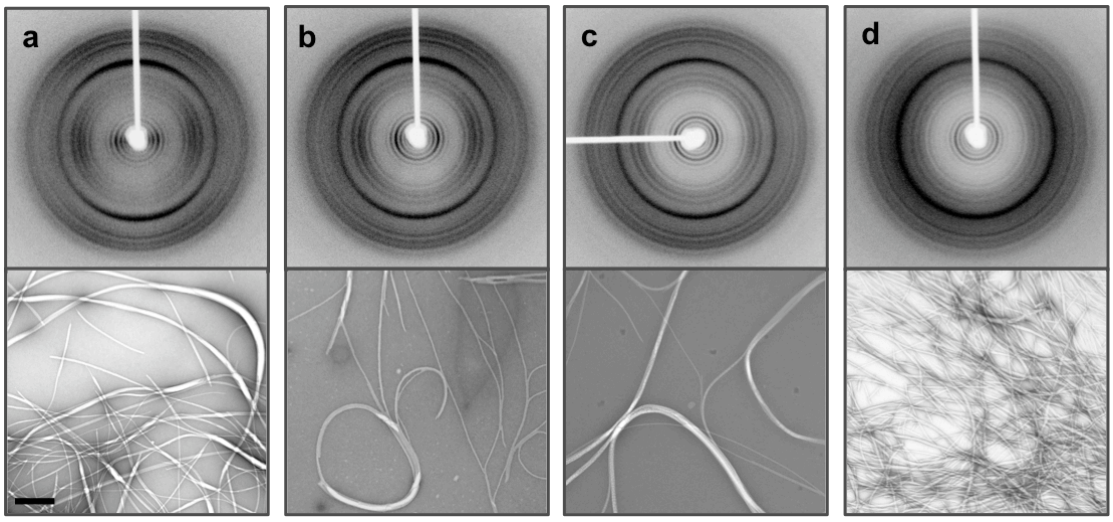


Figure 7.10: Br-AV X-ray fibre diffraction at varied pH 4.7 (a), 4.0 (b), 3.7 (c) and 3.1 (d). Corresponding TEM from targeted hydrogels are shown, scale bar represents 400 nm. All fibre axes are vertical. Figure adapted from (Chen et al., 2010a).

In this manner it was possible to probe the structures equivalent to those formed at 38, 74, 125 and >1360 minutes during time course measurements after addition of GdL (14.42 $mgmL^{-1}$). Figure 7.10 shows the XRFD collected from alignments of fibres of Br-AV targeted at a range of final pHs, the corresponding TEM of fibrils found in the gel matrices at these pHs are also shown. In each case, as expected, fibrillar material is present with comparable morphology to those found over the period of the time course measurements made previously (Figure 7.4) but an appreciable increase in fibril density is observed for the lowest pH of 3.1 (Figure 7.10d).

Comparison of the equatorial and meridional signals, as shown in Figure 7.11, reveals identical traces indicating that no change occurs in the structure of the fibrils across the range of pH environments used for gelation. Some differences are found in the low-resolution regions (data not shown) but this is attributed to variable signal occlusion by the backstop in the XRFD patterns shown in Figure 7.10a-d.

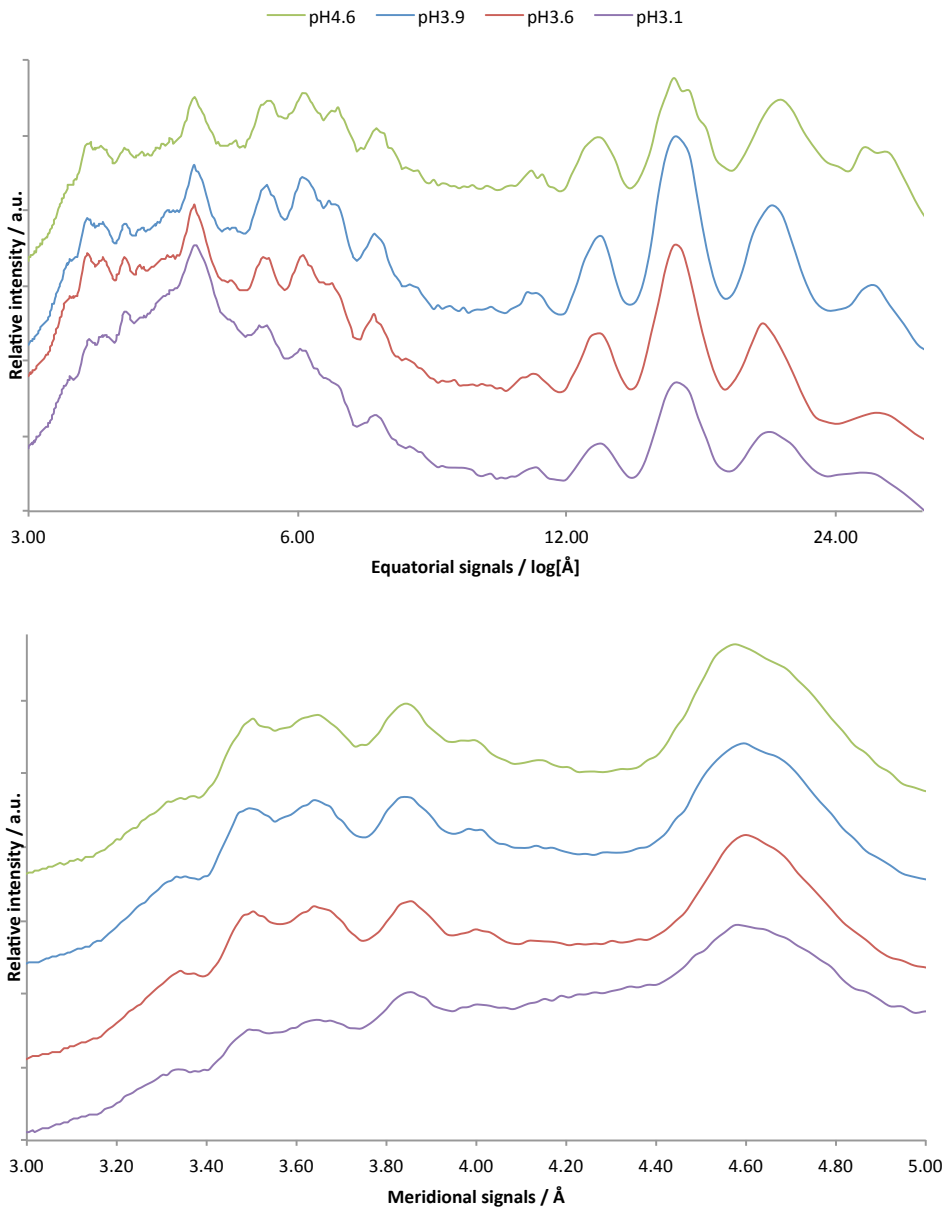


Figure 7.11: One-dimensional intensity traces of the equatorial and meridional signals from a range of pH targeted Br-AV gel fibres. Intensity values have been shifted. To avoid effects from signal occlusion by the backstop equatorial signals are shown only up to 30 Å, where these are over a greater range than meridional signals, a logarithmic scale has been used for clarity. Figure adapted from (Chen et al., 2010a).

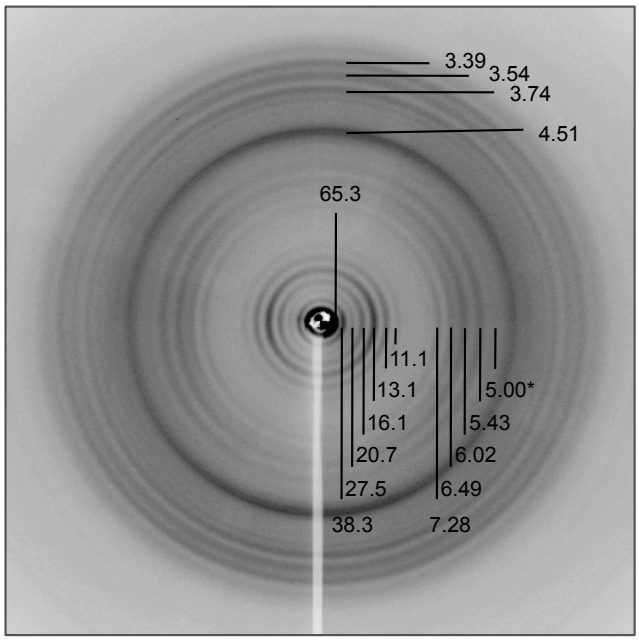
7.2.7 The local structural architecture of Br-AV assemblies

Knowing that the architecture of the fibrils is established rapidly and remains constant over self-assembly and gelation, the final local structural architecture of the fibrils formed by the self-assembly Br-AV was probed by aligning fibrils of a solution of Br-AV (5 $mgmL^{-1}$) after the addition of GdL (14.42 $mgmL^{-1}$).

X-ray fibre diffraction

A synchrotron microfocus beamline (I24, Diamond, Oxfordshire) was used to collect high quality fibre diffraction data from an *in situ* alignment of Br-AV and is shown in Figure 7.12. The pattern is reminiscent of the cross-ß signature for amyloids (Jahn et al., 2010) with a major meridional reflection at 4.51 Å (±0.03) indicating an equivalent repetitive spacing along the fibre axis. The major meridional reflection is compatible with a hydrogen bonding distance and is likely to arise from the inter-ß-strand spacing along the fibre axis. A similar ß-strand spacing of approximately 4.3 Å has been assigned to other dipeptide systems (Smith et al., 2008). The large number of equatorial reflections indicates a remarkably high degree of order perpendicular to the fibre axis. It is also clear that assemblies of Br-AV have appreciably longer-range lateral order than most self-assembled fibrillar systems.

Although all of the reflection positions in the fibre diffraction pattern arise from the dimensions of the repeating lattice, their intensity is modulated by the interatomic distances within that repeating lattice. Since the majority of the interatomic distances are directly related to one or more unit cell dimensions, a repetitive interatomic separation of $a_{1/5}$ will reinforce the 5th order reflection from the *a* dimension. Thus, the intense reflections at higher orders of those arising from the unit cell likely represent repetitive spacings within the unit cell.



Br-AV - microfocus X-ray fibre diffraction pattern

d-Spacing (Å)	Intensity	+d (Å)	-d (Å)	Axis
65.27	252.0	14.27	9.93	E
38.34	127.8	2.39	2.12	E
27.49	91.3	1.98	1.73	E
20.70	102.9	0.61	0.57	E
16.11	123.2	0.33	0.32	E
13.12	98.2	0.18	0.18	E
11.10	82.2	0.16	0.15	E
7.28	90.5	0.05	0.05	E
6.49	96.7	0.06	0.06	E
6.02	104.6	0.04	0.04	E
5.43	103.6	0.03	0.03	E
4.51	122.5	0.03	0.03	M
3.74	100.7	0.02	0.02	M
3.54	96.4	0.02	0.02	M
3.39	93.9	0.02	0.02	M

Figure 7.12: X-ray fibre diffraction from an in situ alignment of Br-AV (5 mgmL^{-1}) assembled by GdL addition (14.42 mgmL^{-1}). Equatorial reflections are indicated by vertical lines and meridional by horizontal lines. Measurements are made in Å using Clearer and a search width of 8 px. Data collected at I24 Diamond, Oxfordshire. The fibre axis is vertical.

The major equatorials are measured to be 65.3 (+14.3/-9.9), 38.3 (+2.4/-2.1) and 16.1 Å (+0.33/-0.32). The low angle reflections likely relate to the large repeating lattice (63.3 & 38.3 Å) of this system but the high angle to the repetitive spacings within the cell. It is interesting that Br-AV modelled in the ß-strand conformation has a length comparable to the 16.1 Å equatorial signal as modelled in Figure 7.13a. It has already been discussed that parallel and antiparallel interpretations from FTIR are controversial (see Chapter 5) however to be able to feasibly explore possible molecular architectures, the structural modelling was limited to antiparallel dipeptide configurations. Considering the antiparallel ß-sheet formation and the fibre axis strand separation of 4.51 Å a limited number of possible ß-strand stacking arrangements can be envisaged as shown in Figure 7.13b-c. These include interamide only (IAO) and interamide-carboxyl (IAC) hydrogen bonding networks that could encourage one-dimensional fibre growth. The bonding scheme including only interamide hydrogen bonding arranges antiparallel dipeptide pairs into a tape with a width of 21.9 Å and the architecture with mixed hydrogen bonding has a tape width of 25.4 Å. It is worth noting that the two widths may

be compared to the equatorial reflections of 20.70 Å (+0.61/-0.57) and 27.49 Å (+1.98/-1.73) respectively.

With some constraint placed on the stacking of adjacent ß-strands, the possible packing of the antiparallel dipeptide pairs perpendicular to the fibre axis was explored. Without considering the naphthalene groups, the lateral interactions may be governed by efficient steric packing and should be relatively weak compared to the directional hydrogen bonding that results in anisotropic fibre growth. Efficient side chain packing of each stacking arrangement was found at ~6 Å corresponding to the equatorial reflection of 6.02 Å. Assuming that the equatorial reflection of 5.43 Å may also represent lateral ß-sheet packing, this was explored as a possible packing distance but was found to cause inter-ß-strand clashing (Van de Waals overlap >= 0.5 Å assessed by Chimera (Pettersen et al., 2004)) (see Appendix II-i).

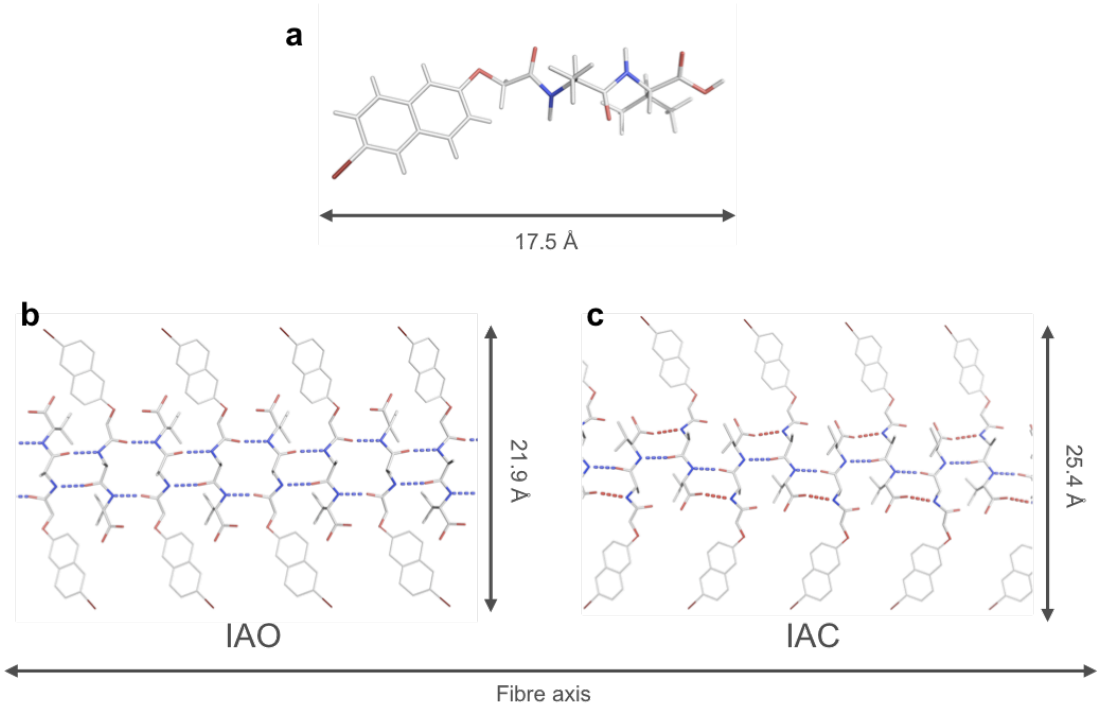


Figure 7.13: Possible Br-AV stacking modes. The monomer of the dipeptide Br-AV with the dipeptide chain in the antiparallel ß-strand conformation (a) and two possible antiparallel hydrogen bonding schemes involving only interamide only (b) or additional interamide-carboxyl hydrogen bonding (c) with inter-ß-strand separations of 4.51 Å. Naphthalene groups are shown for illustration only. Graphics generated in PyMol (DeLano, 2002).

Although the antiparallel ß-strand stacking and packing can be described with some confidence based on experimental evidence, discriminating the naphthalene group architecture and π-π stacking direction is more challenging where freedom exists around the ether linkage. The π-π stacking interaction may extend parallel to the fibre axis and so the naphthyl-naphthyl separation would be governed by the backbone separation of 4.51 Å. In other reports the π-π stacking distance is assigned as being ~3.7 Å based on the interpretation of a WAXS peak at this distance (Hughes et al., 2011, Xu et al., 2010). However, granted by the directional information of the fibre diffraction here it is apparent that this reflection is meridional thus represents an interatomic distance that is aligned to the fibre axis. Where the inter-ß-strand separation is 4.51 Å the naphthyl groups cannot be stacking on the same axis at a shorter distance. The 3.74 Å reflection here, is more likely to be a higher order reflection from the lattice dimensions.

To most accurately model the π-π interaction direction and geometry the exciton coupling model, already described in Figure 7.9, applicable to the CD and UV/Vis data was explored. Further, assuming the exciton coupling model is applicable to this system, comparison of LD measurements to the CD were able to determine that the naphthalene long axes lie between 0 – 54.7° relative to the fibre axis (Chen et al., 2010a) (see Appendix II-ii). To construct the most representative model of Br-AV in the fibrillar phase these constraints must first be considered in the context of the lattice they stabilise in.

X-ray fibre diffraction pattern indexing and unit cell determination

A unit cell cannot be smaller than the lowest resolution/largest reflection measured on the XRFD pattern. As the single major meridional signal arises from a unique axis it can be assigned to the dimension parallel to the fibre axis, here assigned as *b*. Pattern indexing then followed to explore the possible *a* and *c* dimensions to find potential unit cells. These searches generally resulted in large unit cell solutions (see Appendix II-iii). Early searches did not include the 65.27 Å (+14.27/-9.93 Å) equatorial reflection where this was later collected using a

synchrotron beamline. Although these limited searches produce the best mathematical indexing of the fibre diffraction some predicted reflections lie outside of the systematic error associated with the measured experimental signal position measurement (see Appendix II-iii).

The cell that theoretically reproduces all of the experimental reflections within the systematic errors is considered the best cell determination as shown in Table 7.3. Simulated fibre diffraction from the unit cell lattice excluding a structure confirm the cell shown in Table 7.3 best reproduces the low resolution reflections of the experimental fibre diffraction pattern (see Appendix II-iv). Despite this, however, the cell is still large relative to the dipeptide molecule allowing for a great number of packing arrangements, making *de novo* modelling problematic. Initial attempts were made to arrange individual dipeptide molecules within the indexed unit cell, but were unable to reproduce the fibre diffraction signals observed in the experimental pattern (see Appendix II-v). Thus, only the minimum asymmetric unit was modelled based on the experimental constraints.

BrAV unit cell					
Dimension	**Search limit**	**Prediction**	**Observed (Å)**	**[h k l]**	**Predicted (Å)**
a (Å)	40	37.29	65	[0 0 1]	72.45
b (Å)	9.02	9.02	38	[1 0 0]	37.29
c (Å)	80	72.45	27	[1 0 2]	25.98
α (°)	90	90	21	[1 0 3]	20.27
ß (°)	90	90	16	[1 0 4]	16.29
γ (°)	90	90	13	[2 0 4]	12.99
			11	[3 0 3]	11.05
Error (%):	0.0262		7.3	[5 0 2]	7.3
			6.5	[4 0 8]	6.5
			6.02	[6 0 3]	6.02
			5.43	[3 0 12]	5.43
			5	[7 0 5]	5
			4.51	[0 2 0]	4.51

Table 7.3: Putative unit cell indexing of the Br-AV pattern.

7.2.8 Br-AV modelling

Using the determined unit cell dimensions, the constraints on ß-sheet architectures and naphthalene geometry and orientation, models of Br-AV in the fibrillar state were constructed. Both the IAO and IAC packing modes may be modelled into the unit cell where the long axis of the dimer is aligned to the *a* dimension of the determined unit cell. In both the IAO and IAC arrangements the naphthalene groups of adjacent dimers overlap into a stacking arrangement. Assuming the 37.3 Å *a* cell dimension, due to the shorter long axis of the IAO dimer the naphthalene groups are displaced relative to one another when viewed down the fibre axis (see Appendix II-vi). The IAO model at initial visual inspection may be more favourable where this displacement produces non-ideal face-to-face arrangement and rationalises the small hypsochromic stokes shift of the B_b transition upon assembly. With the IAO and IAC antiparallel ß-strand stacking modelled in the unit cell the naphthalene orientation and relative geometries were explored within the spectroscopic constraints by manual adjustment of the torsion angles α and ß around the ether linkage (Nap-**α**-**O**-**ß**-Ala-Val). No obvious configuration could be found for the IAC architecture by this manual search. However in the antiparallel IAO architecture the α and ß torsion angles of the ether group were adjusted to -173° and -139° respectively as shown in Figure 7.14. In this configuration, exciton coupling is presumed to be possible through all dimensions of the lattice and on inspection almost all of the orientation and geometrical constraints regarding the naphthalene group architecture are satisfied. The modelling would benefit from a thorough computational search through the ether bond torsion angles to find the best naphthalene architecture, and entirely rule out the IAC stacking mode. Nevertheless, this satisfies most of the constraints placed on the naphthalene geometries by spectroscopic data. The naphthyl long axes are orientated at ~55° relative to the fibre axis (a); the nearest neighbour B_b naphthyl transitions are face-to-face but with an α value of ~76° (b) accounting for the small hypsochromic shifting but created by the H-aggregate geometry with a θ value of ~80° (c). The L_a transitions are oblique although need geometrical rearrangement where their θ value is ~76° (d).

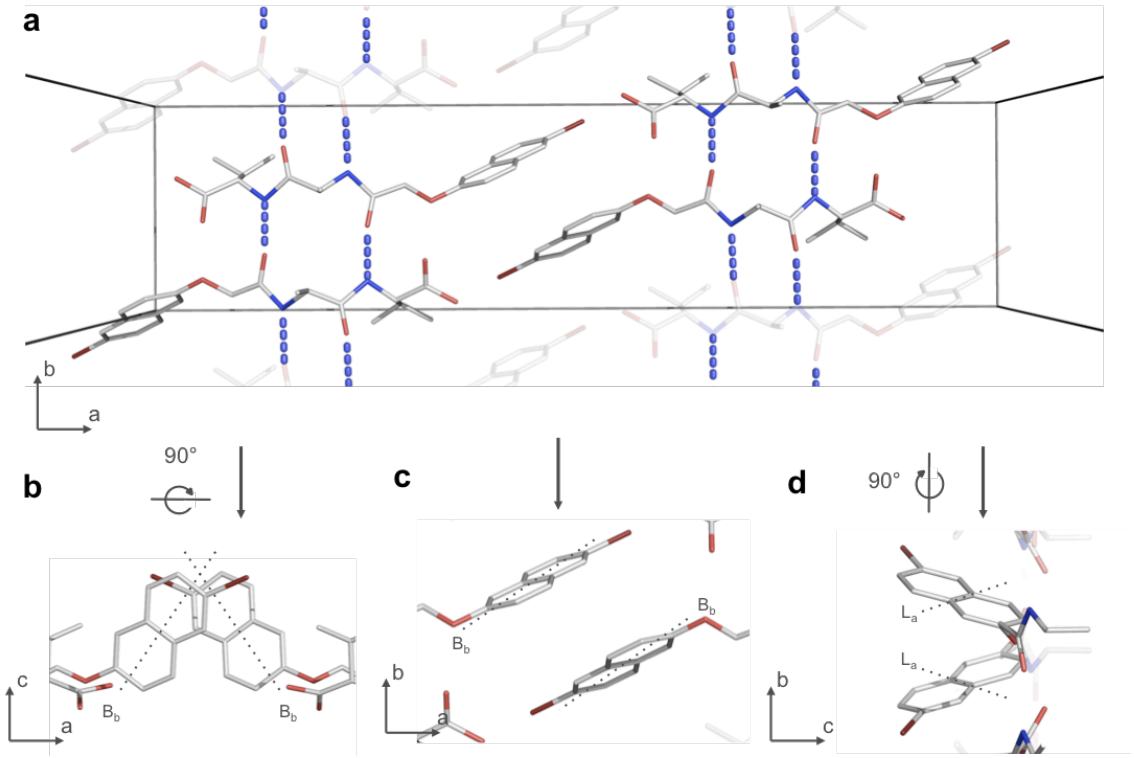


Figure 7.14: A preliminary model of the structure of Br-AV in the fibrillar state representing the essential interactions that stabilise the unit cell. ß-strands are found in an antiparallel arrangement, separated by 4.51 Å and stabilised by interamide hydrogen only (blue) (a). ß-sheets are separated perpendicular to the fibre axis by 6.02 Å for efficient side chain packing. The α and ß torsion angles of the ether linkage are ca. -173° and 139° approximately orientating the naphthyl groups relative to the fibre axis and neighbours consistent with the spectroscopic data (b-d). High-energy long axis and low-energy short axis naphthyl transitions are labelled with B_b and L_a respectively. Graphics generated in PyMol (DeLano, 2002).

Br-AV model unit cell packing and X-ray fibre diffraction simulation

Assuming the interatomic spacings within the structure will be related to the unit cell dimensions, the interatomic spacings of the asymmetric unit can be used to create the interatomic spacings of the cell through the use of symmetry and translation functions. If the diffraction pattern, in particular the equatorial reflections, were only governed by the repeating lattice then the intensity of the higher order reflections would decrease proportional to the reflection order number. A high intensity reflection, however, indicates a repetitive spacing within the unit cell. The model shown in Figure 7.14 is consistent with the intense meridional 4.51 Å (+0.03/-0.03) reflection and the intense equatorial 38.34 Å (+2.39/-2.12) (2^{nd}), 16.11 Å (+0.33/-0.32) (3^{rd}), and 6.02 Å (+0.04/-0.04) (4^{th}) reflections (reflections ranked by intensity).

Using the packing architecture shown in Figure 7.14, the unit cell can be modelled in the *a* dimension by simple translation through *a/2* and *b/4*. The repetitive packing separation along the *c* axis is modelled at 6.02 Å, but where this is the 12th order reflection of the *c* dimension it is unclear what makes each symmetry mate along this axis unique to fully populate the unit cell. As such, full unit cell modelling in a lattice of this large size is problematic. Thus, the packing architecture in the *c* dimension was modelled as far as a pair of ß-sheets separated by 6.02 Å. Though the model reasonably fits the data, attempts to simulate the full pattern from the modelled structure were unsuccessful (see Appendix II-vii). This is likely in part due to the inability to model the variation along the *c* (72.45 Å) dimension.

As the unit cell cannot be modelled in full, its accuracy cannot be truly verified by simulated X-ray fibre diffraction. The model of the asymmetric unit may be accurate but without the correct cell packing, the validity of the structure cannot be faithfully validated. Where the discrimination here is made between the IAO and IAC architectures based on the predicted unit cell dimensions neither conformation can, at present, be wholly ruled out. That said, the models are based on strong biophysical data and on as many available structural details as are currently interpretable. As such the IAO architecture currently appears to be the most plausible, though this should be treated with caution until more experimental evidence becomes available or more thorough computational modelling approaches can be adopted. This raises the point that models may be proposed and be reasonably based on structural and biophysical evidence (Hughes et al., 2011, Smith et al., 2008, Xu et al., 2010) but they must be validated against the structural data by comparison of simulated data. Nevertheless, of importance here, whether treating the models with some caution or taking the unit cell prediction as accurate, some interesting conclusions and questions can be posed

The question can be raised as to whether the large cell is a true cell or simply the dimensions of the protofilament packing. In this case, perhaps the asymmetric model presented here represents a sub-lattice of a pseudo cell: a concept observed for other self-assembled systems (Sikorski et

al., 2003). If the diffraction pattern is governed by the dimensions of bundles of fibrils this too is interesting and potentially important as it indicates ordered lateral arrangement and macromolecular organisation of fibrils in the gel phase. This, in turn, may possibly influence material properties where macromolecular fibril interactions are known to govern hydrogel behaviour (Peppas et al., 2009). Given the identification of different macromolecular fibril organisations could these be related to material properties?

7.2.9 The material properties of hydrogels from assembly matrices

The gels formed at various pH ranges were found to have complex rheological behaviours. With the exception of the gels targeted for pH values >4.5 across the pH range investigated the final G' and G" values arising from the viscoelastic properties of the gels were extremely similar indicating comparable gel strengths (Chen et al., 2010a). However, other measured rheological properties displayed differences, in particular the gels response to strain. By applying an external force/strain and observing when G' is no longer significantly larger than G" it was possible to measure the "gel-like" to "liquid-like" transition indicative of breakdown of the hydrogel matrices. It was found that the gels formed at higher pH were much less tolerant to strain than those formed at lower pH, where a breakdown of the high and low pH gels was observed respectively at 1% and 5% strain. However the "gel-like" to "liquid-like" transition in the low pH gels occurs rapidly and completely, whereas the high pH gels maintain a G'/G" difference from the initial breakdown at 1% up until 8% strain (Chen et al., 2010a). These data are consistent with the fibrillar matrices across all the pH ranges producing gels of comparable strengths but at high pH the fibrils behave elastically and at low pH are rigid. The elastic fibrils formed at high pH are more readily subject to structural breakdown, although can withstand high amounts of strain presumably granted by their elasticity. The rigid fibrils formed at low pH less readily undergo the "gel-like" to "liquid-like" transition but do so under lower strain (Chen et al., 2010a). The rheological behaviour is also seen in the similarities and differences in the XRFD patterns when targeted to specific pHs. As the pH is targeted to lower values, the radial

averaging exhibited by the fibre diffraction increases (see Figure 7.10). This is consistent with an inability to effectively align the fibrils. This correlates well with the rheological data where at low pH the gel matrices are resistant to structural reorganisation and breakdown or in this case rearrangement into a fibrous alignment.

Whilst the XRFD patterns indicate no local structural change across the range of pHs assemblies are formed at, the differences in material viscoelastic properties appear to be attributable to the organisation of the gel matrices at length scales possible greater than observed here. Whether the two can be connected is addressed in the subsequent Chapter 8.

7.3 Summary

At high pH the dipeptide molecules are non-fibrillar and the solution they are dissolved in has a viscosity comparable to water. Figure 7.15 shows the theoretical ionisation state of the carboxy-terminus and shows that where alkaline conditions are maintained by NaOH the carboxy-terminus will be in the negative carboxylate form. The negative charge of this terminus may prevent close interaction of dipeptide molecules but where at high pH these molecules have been observed to form micelles (Chen et al., 2011) this would be consistent with the carboxylate presenting a hydrophilic face to the water solution whilst the hydrophilic portion of the molecule is buried in the micelle. Upon lowering of the pH below the pKa of the carboxy-terminus, the protonated carboxyl is formed. This group could hydrogen bond with water but the observation of self-association with pH decrease suggests that the stability of these interactions (y_1 – kJ mol^{-1}) are less than the stability of potential inter peptide interactions (y_2 – kJ mol^{-1}) resulting in increasing proximity of dipeptide molecules and self-assembly. Alternative mechanisms of self-assembly and gelation has been reported in a special case where divalent cations are used to induce assembly (Chen et al., 2011).

As the pH continues to lower below the pKa of the carboxy-terminus self-assembly then proceeds through the mechanism illustrated in Figure 7.16. Through the inter-molecular structuring of the dipeptide and naphthalene portions of the molecule. The naphthalene groups enter into an ordered chiral environment, in close enough proximity to become exciton coupled. The dipeptide simultaneously rapidly adopts (0 – 40 mins) a ß-sheet fibrillar conformation, forming before gel strength is fully acquired.

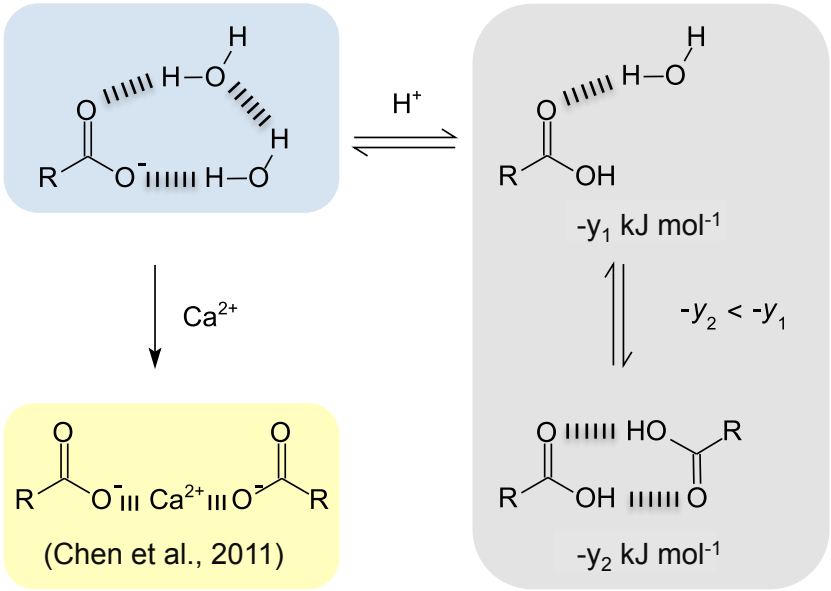


Figure 7.15: Hypothetical ionisation states of the carboxy-terminus of a dipeptide. At high pH where the terminus will be in the carboxylate form, hydration may favour micelle formation (blue). Protonation of the carboxy-terminus by lowering of pH may result in self-association interactions ($-y_2$) being more energetically stable than hydrating interactions ($-y_1$), subsequently leading to gelation (grey). In alternative mechanisms the addition of divalent cations has also been reported to induce gelation (yellow) (Chen et al., 2011). Three lines indicate ionic interactions and six lines hydrogen bonding.

The ordering of the naphthalene groups essentially occurs concurrently with the structuring of the dipeptide backbone, thus neither is obviously the principle driving force for assembly. The high hydrophobicity of the naphthalene group and dipeptide chain may promote the hydrophobic collapse of the molecules in solution. However, at high pH the carboxyl groups would exist in a charged carboxylate form, as shown in Figure 7.15, resulting in electrostatic repulsion of neighbouring molecules (Figure 7.16) but also favourable water solvation energies stabilising micelle structures (Chen et al., 2011). As the carboxyl group is protonated, self-assembly can occur via hydrogen bonding and π-π interactions. It is interesting that π-π stacking interactions are not significantly perturbed by the 6-position bromine atom on the naphthalene, which would be expected to be inductive drawing electrons away from the delocalised π-electron orbitals.

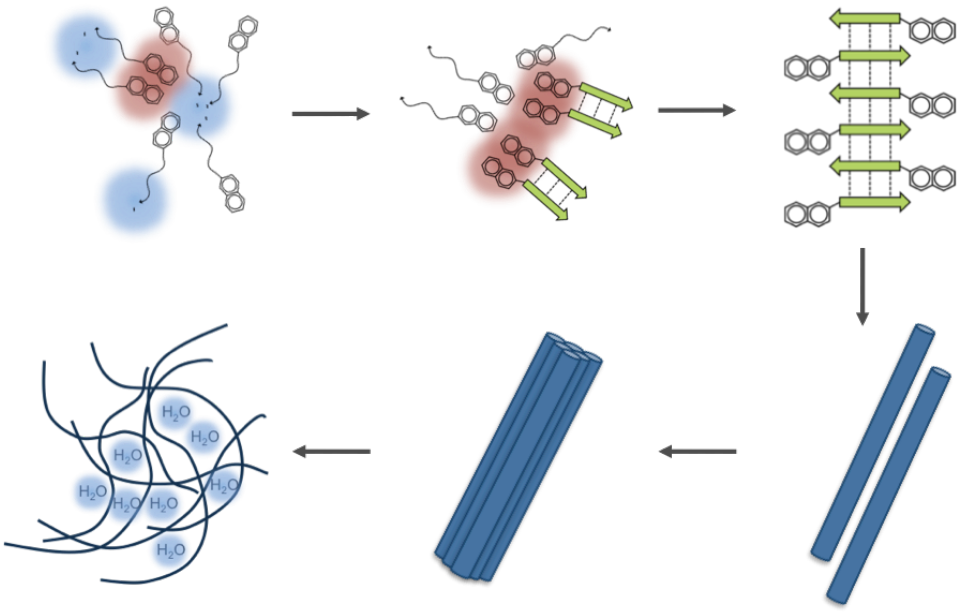


Figure 7.16: An illustrative summary of the assembly mechanism/pathyway and gelation of Br-AV.

As the pH continues to fall, the local structure of the fibres remains unchanged but an increase in the lateral association, fibril crosslinking and entanglement in the gel matrix contributes to the developing gel strength. The gel strength is stable throughout the assembly process but the fibrillar matrices formed at early time points have elastic characteristics. However, as pH continues to fall, the resistance of the hydrogel to deformation is increased as the fibrils become more rigid, possibly due to fibril lateral association.

The indexing of the fibre diffraction pattern to large unit cells makes full structural solutions difficult to ascertain due to large degrees of freedom in the possible packing arrangements within the cells, though this is addressed in Chapter 8. On the short-range Br-AV molecules assemble into a cross-ß like structure, likely in an interamide only hydrogen bonding pattern with antiparallel ß-strands separated by 4.51 Å along the fibre axis. The ß-sheet separation is modelled at 6.02 Å and in this arrangement the naphthalene groups are able to interact in a face-to-face manner with regard to their long-axes whilst their short axes are at an oblique end-to-end arrangement.

Although the short-range structural architecture of the fibrils can be modelled, the lateral association of protofilaments into fibres is difficult to describe accurately through modelling due to high degrees of freedom in potential packing arrangements. Nonetheless, the long-range packing arrangements of fibrils is extremely ordered. Gel strength appears to not be directly derived from the local dipeptide architecture but from the structural organisations of the fibrils themselves and consequently their involvement in the gel matrix. Where the local architecture of the fibres is unchanged over the development of gel strength, the link between structures on the scales of up to ~70 Å and the properties of the fibril interactions, and thus gel strength, is not entirely clear. Solid-state NMR and cryo-EM studies may greatly benefit establishing this connection.

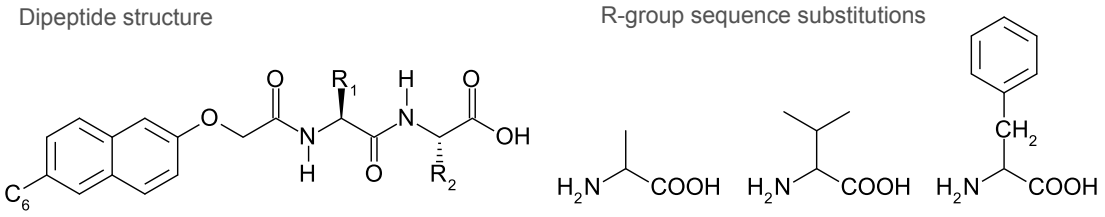
8 Results and Discussion: The structure, assembly and gelation of a library of low-molecular weight gelators

8.1 Introduction

The link between molecular structure, self-assembly and bulk material properties remains unclear. It is imperative to understand the assembled structures of low-molecular weight gelators (LMWGs) that give rise to gelation to subsequently understand why one gelator produces a material with different properties to another. Structural determination of the fibrillar phase has to date been problematic, not allowing a direct comparison of structural information to material property information. One approach by which to address this is to use single crystal or powder X-ray diffraction structure data and to assume these structures represent the fibrillar phase (Dastidar, 2008), however this assumption is a point of contention where the crystal phase may not represent the fibre phase (Houton et al., 2012).

8.1.1 Naphthalene conjugated dipeptide library

A library of twelve naphthalene conjugated dipeptides was created to explore the relationship between molecular structure, self-assembly, fibrillar structure and hydrogel properties. Naphthalene conjugates included 2-naphthol, 6-bromo-2-naphthol and 6-cyano-2-naphthol. The dipeptide sequence was varied between aliphatic and aromatic sequences but all were hydrophobic, including Ala-Ala, Ala-Val, Phe-Phe and Phe-Val. Table 8.1 shows the dipeptide library where three positions were varied to produce the library: the atom at the 6-position on the naphthalene ring, referred to as C_6 and the first and second amino acid residue, referred to as R_1 and R_2 respectively.



#	Dipeptide	C_6	R_1	R_2	Apparent pKa†	clogP†	Dipeptide hydropathy index*	Phase†	G' / Pa†	G'' / Pa†
1	2-AA	H	L-Ala	L-Ala	5.1	-0.16	3.6	Gel/Crystal[1]	83000	931
2	Br-AA	Br	L-Ala	L-Ala	4.9	0.63	3.6	Gel	28	10
3	CN-AA	CN	L-Ala	L-Ala	5	-0.43	3.6	Turbid Gel	79000	7800
4	2-AV	H	L-Ala	L-Val	4.2	0.62	6	Gel/Crystal[2]	n.d.	n.d.
5	Br-AV	Br	L-Ala	L-Val	5.8	1.4	6	Gel	25000	560
6	CN-AV	CN	L-Ala	L-Val	5.8	0.35	6	Gel/Crystal[3]	n.d.	n.d.
7	2-FF	H	L-Phe	L-Phe	6	2.76	5.6	Gel	72000	1200
8	Br-FF	Br	L-Phe	L-Phe	6.8	3.55	5.6	Gel	180000	700
9	CN-FF	CN	L-Phe	L-Phe	6.6	2.5	5.6	Gel	15000	280
10	2-FV	H	L-Phe	L-Val	6.5	2.08	7	Turbid Gel	2000000	11000
11	Br-FV	Br	L-Phe	L-Val	6.6	2.87	7	Turbid Gel	100	1
12	CN-FV	CN	L-Phe	L-Val	6.1	1.81	7	Turbid Gel[4]	29000	580

*Table 8.1: The twelve naphthalene conjugated dipeptide systems of the library, varied by the naphthalene conjugate and the amino acid sequence of the dipeptide. †Apparent pKa and clogP values are in water and phase formation as reported in (Chen et al., 2010c). *Hydropathic index calculated from the dipeptide sequence only using the Kyte and Doolittle scale (Kyte and Doolittle, 1982). [1]Gelates and then with time crystallises in the same solvent (H_2O). [2]Directly crystallises from solution, under very specific conditions a gel is transiently formed. [3]Reported to form a crystal phase, in these experiments observed to enter the gel phase. [4]Reported to form a turbid gel, in these experiments observed to enter the gel phase.*

Table 8.1 makes reference to naphthalene-conjugated dipeptides where the peptidic portion is conjugated at the 2-position of the naphthyl ring, however further various LMWG systems were created by conjugating to the 1-position. Conjugating the dipeptide sequences, Gly-Ala and Ala-Gly (subsequently referred to as 1-GA and 1-AG), to 1-Naphthol two further dipeptide systems were created as previously reported (Adams et al., 2010a).

8.1.2 Experimental Aims

By structurally and biophysically characterising a library of dipeptides this work aims to understand the link between the dipeptide molecule identity (sequence and naphthyl conjugation), molecular fibrillar architecture and material properties. The link between dipeptide molecular structure and material properties is anticipated to be a complex relationship, as described in Chapter 7, but the use of a library of dipeptides allows an opportunity to examine the similarities and differences between dipeptide molecular structure such that they

might be linked to material properties. A comparison between the fibre and crystalline phase is also made to attempt to reveal the similarities and differences between these two states. Attempts are made to find a relationship between self-assembly structure and gel material properties revealing the complexity of this link but disseminating the distinction that must be made between the two phenomena. By systematic variation of the dipeptide sequence between aliphatic and aromatic sequences, the roles of peptidic interactions as well as π-π stacking interaction may be understood. An understanding of this may lead to sequence design rules for LWMG design. Systematic substitution of the 6-position of the naphthyl group between Br and CN groups allows the investigation of the role of naphthyl π-π stacking interactions which have great emphasis placed on them in the literature (Jayawarna et al., 2006, Mahler et al., 2006, Mart et al., 2006, Toledano et al., 2006, Xu et al., 2007, Yang and Xu, 2004, Zhou et al., 2009).

This chapter consists of mainly unpublished data but builds upon the material analysis performed by Chen *et al.* in (Chen et al., 2010c). Where a comparison of the structure in the fibre phase to the structure of the crystal phase is informative, the analysis performed by Adams *et al.* is also expanded upon (Adams et al., 2010a). Dr. D. J. Adams (University of Liverpool) is gratefully acknowledged for synthesising and providing the dipeptide materials. Daniela Hochegger (University of Sussex) is acknowledged for assistance in collection of circular dichroism data for the dipeptide library. Stephen Thompson (University of Sussex) is gratefully acknowledged for assistance in alignment and X-ray fibre diffraction pattern collection for the systems CN-FF and CN-FV. Dr. D. J. Adams (University of Liverpool) is acknowledged for providing the atomic co-ordinates solved by Dr. M. Schmidtmann (University of Liverpool) of the dipeptide 2-AV, Dr. J. Bacsa (University of Liverpool) of 1-GA and 1-AG and Dr. H. Sato and Dr. A. Yamano (Rigaku) of 2-AA in the crystalline phase. Dr. L. Chen (University of Liverpool) is gratefully acknowledged for collecting and providing the Fourier transform infrared spectroscopy data for the dipeptide library in the assembled state. The comparison of the 2-AA fibre to the crystal phase is reported in a manuscript under review (Houton et al., 2012).

8.2 Results and Discussion

8.2.1 Pre-gelation general observations

General observations

The stock solutions at high pH were observed to have a viscosity visually comparable to water, except for diphenylalanine containing dipeptides. The systems 2-FF, Br-FF and CN-FF were found to have a viscosity discernibly above water when subject to pipetting. All dipeptide solutions failed a vial inversion test except CN-FF, which was found to be self-supporting (data not shown). The cause of the greater viscosity of the FF containing dipeptides is attributable to the presence of pre-assembled ordered material. This is supported by the birefringence exhibited by dried drops of stock solutions when observed by cross-polarised microscopy as shown in Figure 8.1. However, it is important to note that although 2-FF, Br-FF and CN-FF were observed to be viscous, they are not considered to be gels as G' is not larger than G" (Dr. D. J. Adams, verbal communication).

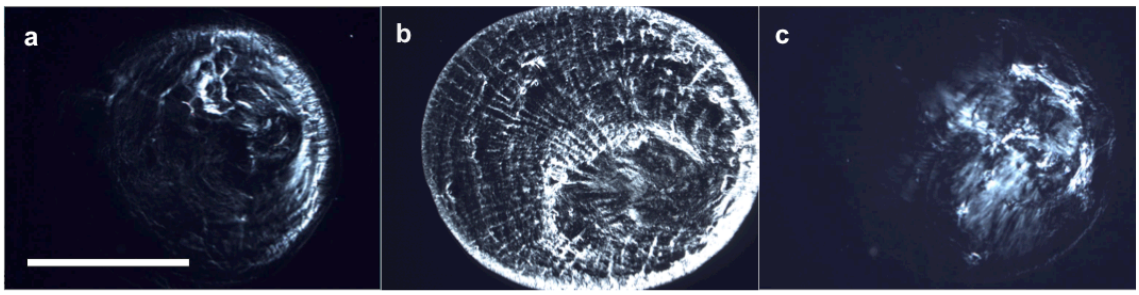


Figure 8.1: Cross-polarised micrographs of dried samples of 2-FF (a), Br-FF (b) and CN-FF (c) at pH 10.5. The scale bar represents 5 mm.

Spectroscopic observations

To determine the structure of the dipeptides at high pH (pH 10.5), biophysical characterization was carried out. To best understand any results generated from this, a full UV/Vis characterization of the dipeptides was initially performed. Three chromophores can be considered to contribute to the absorbance characteristics of the naphthalene conjugated dipeptide molecule; naphthalene, the two amide bonds of the peptide backbone and, where present, phenylalanine. Table 8.2 shows the positions and expected strengths of each electronic transition. From this basic analysis of the excitable electronic orbitals in a naphthalene dipeptide system, it can be seen that the near-UV absorption bands of naphthalene are almost distinct from the far-UV higher energy transitions associated with the peptidic portion of the molecule. It should be considered that the electronic excitability of these chromophores are sensitive to substitution, environment (see Chapter 3), and thus structure. The naphthalene group is particularly sensitive to environment and substitution (see Chapter 7), however despite this the electronic transitions between ~180 – 221 nm can be approximated to arise from the dipeptide backbone and those between ~231 – 312 nm to arise from the naphthalene chromophore with the exception of the weak L_b transition of phenylalanine.

Chromophore	Assignment	Excitation / nm	ε_{max}
Phe	B_b	180	60,000
ß-sheet	π-π*	195	
Phe	L_a	210	10,000
ß-sheet	n-π*	219	
ß-sheet	n-π*	221	
Nap	B_b	231	
Phe	L_b	260	200
Nap	L_a	278	
Nap	L_b	312	

Table 8.2: The electronic transitions of the amide bond in the ß-sheet conformation (Woody, 1996), phenylalanine (Woody, 2007) and naphthalene (Singh and Thakur, 1981). Where the extinction co-efficient of individual transitions are reported in the literature, these are shown.

The UV/Vis characterisation of pre-gelation dipeptide stocks at pH 10.5 are shown in Figure 8.2 and are clearly dominated by the absorption bands of naphthalene. From the intensities and relative positions of the absorption bands from each dipeptide it is clear that the naphthalene electronic transitions are most directly affected by the nature of the 6-position conjugation; either H (unsubstituted), Br or CN. However, substitution at the 2-position has also been shown to perturb the expected electronic transitions of naphthalene, in some instances by as much as 12 nm (Singh and Thakur, 1981). In respect of this, the direct comparison of the spectral properties of the substituted naphthalene groups should be done so with care, although it is still possible. Table 8.3 shows the comparison of Ala-Ala containing dipeptide systems to comparable 6-position substituted naphthyl groups. A close correlation can be found between the dipeptide system and its equivalent 6-substituted-2-naphthol. This characterisation is useful for identifying the dipeptide system and means that the dipeptides can be grouped spectrally by the nature of their naphthalene conjugate. Table 8.3 shows the assignment of these absorption bands to particular electronic transitions across the entire library. The B_b transitions are most simply assigned as they are expected to have the strongest absorbance; the L_a and L_b transitions are more tentatively assigned.

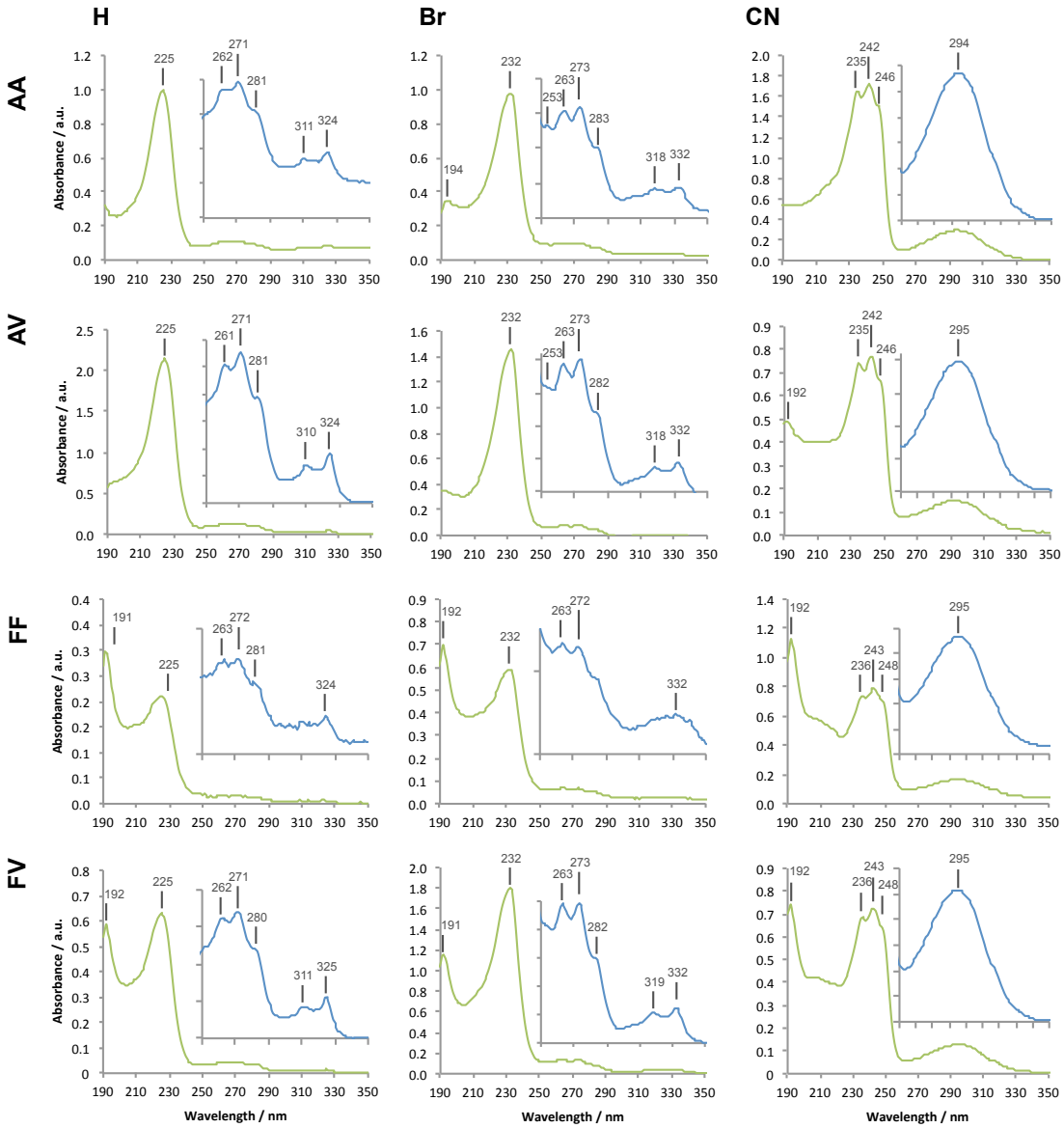


Figure 8.2: The UV/Vis absorbance of aliphatic and aromatic naphthalene dipeptide stocks at pH 10.5 at 6.25 (green) and 25.0 µgmL^{-1} (blue).

	Absorption bands / nm			
Assignment	**2-Naphthol**	**2-AA**	**6-Bromo-2-Naphthol**	**Br-AA**
B_b	224	225	232	232
		262		253
		271		263
L_a	278	281	276	273
				283
		311		318
L_b	324	324	339	332

Table 8.3: A comparison of the electronic transitions exhibited by the dipeptide 2-AA to 2-hydroxyl-naphthalene (MO calculations) (Singh and Thakur, 1981), Br-AA to 6-Bromo-2-Naphthol (in acetonitrile) (Dempsey et al., 2010). No comparable data was found for an appropriate 6-Cyano-2-Naphthol.

Across the range of dipeptides, it is noticeable that the presence of the FF motif largely attenuates the absorbance of the near-UV naphthalene transitions, suggesting a possible interaction of the electronic structures of both these aromatic groups prior to gelation. Additionally, the decreased absorbance from the naphthalene groups of phenylalanine containing dipeptides may also be an indication that the naphthalene groups are in close association with one another in an assembled fibre prior to gelation. These data are consistent with the previous observations of ordered material in the stocks of diphenylalanine containing dipeptides. It is also apparent that the conjugation of a cyano group at the 6-position entirely perturbs the near-UV naphthalene transitions.

Nap			Br-Nap			CN-Nap		
λ / nm	**±λ / nm**	**Assignment**	**λ / nm**	**±λ / nm**	**Assignment**	**λ / nm**	**±λ / nm**	**Assignment**
191.5	0.7		192.3	1.5		192	0.0	
225	**0.0**	**Bb**	**232**	**0.0**	**Bb**	**235.5**	**0.6**	**Bb**
						242.5	0.6	
			253	0.0		247	1.2	
262	0.8		263	0.0				
271.3	0.5		272.8	0.5				
280.8	**0.5**	**La**	**282.3**	**0.6**	**La**	294.75	0.5	
310.7	0.6		318.3	0.6				
324.3	**0.5**	**Lb**	**332**	**0.0**	**Lb**			

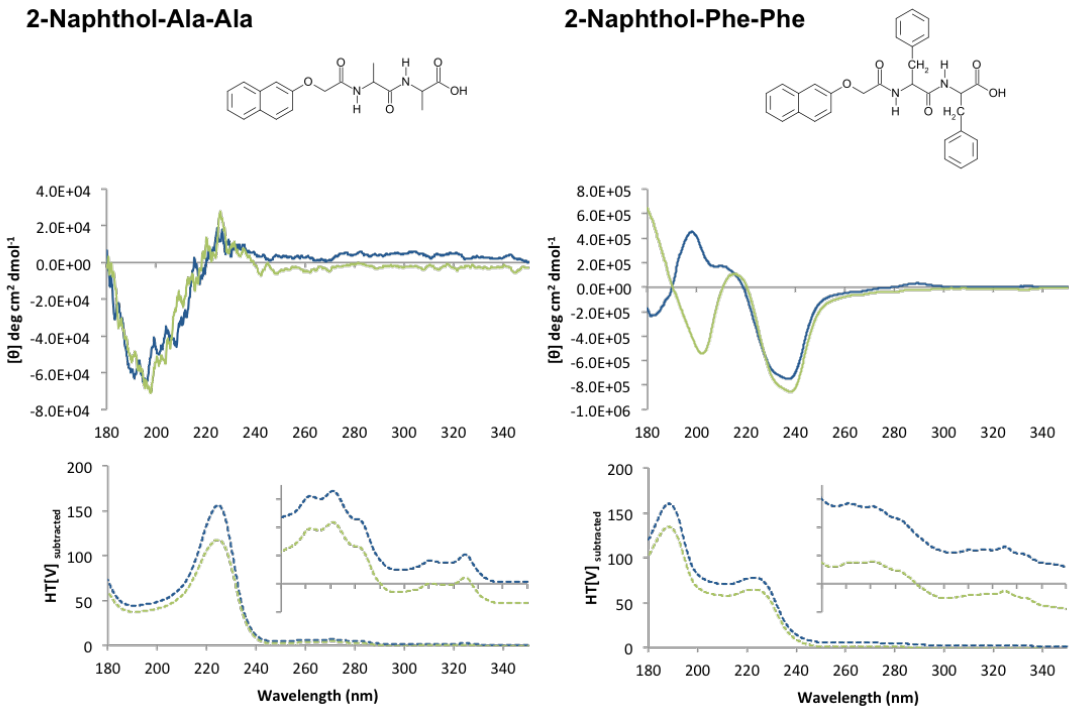
Table 8.4: Assignment of the absorption bands of each naphthalene dipeptide derivative to electronic transitions B_b, L_a or L_b, showing very little variation across the library between naphthalene related systems.

With an understanding of the electronic excitability of the dipeptide systems, circular dichroism (CD) was then used to determine the structure of the dipeptides at high pH (pH 10.5). The full CD characterisation of the dipeptide stocks are shown in Appendix III-i. Each of the aliphatic dipeptide systems are found to have a CD spectra consistent with a random coil structure with, as expected, no orientation dependency in the signals. The FV systems show no signals attributable to pre-gelation structure, but further corroborating the visual and UV/Vis observations, the diphenylalanine containing dipeptides are found to be structured and exhibit orientation signal dependence inferring anisotropic assembly pre-gelation. Where these systems all have apparent pKa values below 6.8 (see Table 8.1) according to the current understanding of the protonation mechanism of assembly (see Section 2.4.4), self-assembly giving rise to gelation will not proceed. The LMWG at high pH might be thought of as being monomeric but likely adopt some structure; it has been reported that Br-AV at high pH forms micelle-like structures (Chen et al., 2011) but certainly these structures, as measured here, do not possess the anisotropy associated with self-assembled fibres.

Figure 8.3 shows the CD from an un-substituted naphthalene conjugated dipeptides with the sequence AA and FF, illustrating the trends in the whole library. 2-AA exhibits a random coil dipeptide with an $HT[V]_{sub}$ profile characteristic of monomeric 2-AA. 2-FF exhibits a CD spectra that is not easily interpretable within the protein CD spectroscopy definitions but the presence of signals indicate structured assemblies that are anisotropic due to the directional dependence of signals, consistent with fibrillar assemblies. The B_b transition of the naphthalene is also attenuated by approximately 50.1 %, relative to 2-AA, indicating the close association of these groups pre-gelation (see Appendix III-ii).

As discussed, however, at high pH the dipeptides are not gels despite the clear evidence of pre-assembly of diphenylalanine dipeptides. This may indicate that self-assembly should be considered distinct from gelation, as discussed later. To investigate the structure of the gel phase

Glucono-δ-Lactone (GdL) was used to reproducibly lower the pH of solutions below the pKa values of the carboxy termini of the dipeptides (Adams et al., 2009).



CD

0deg		90deg	
nm	[θ] deg cm² dmol⁻¹	nm	[θ] deg cm² dmol⁻¹
195.7	-4.55E+04	197.6	-4.93E+04
224.6	1.31E+04	225.9	1.95E+04

HT[V]

0deg		90deg	
nm	HT[V] subtracted	nm	HT[V] subtracted
224.7	155.84	224.5	117.46

CD

0deg		90deg	
nm	[θ] deg cm² dmol⁻¹	nm	[θ] deg cm² dmol⁻¹
334.2	8.41E+03	332.4	-2.13E+04
		319.3	-1.71E+04
289.2	3.17E+04	291.0	-2.28E+04
237.2	-7.49E+05	238.2	-8.55E+05
210.1	1.72E+05	215.1	1.11E+05
198.2	4.56E+05	202.0	-5.45E+05
182.3	-2.35E+05	<180.0	>6.42E+05

HT[V]

0deg		90deg	
nm	HT[V] subtracted	nm	HT[V] subtracted
188.3	160.20	188.1	134.38
222.8	77.84	222.7	64.53

Figure 8.3: Examples of the circular dichroism exhibited by aliphatic dipeptide systems at high pH, have no aromatic interaction and a random coil peptide structure (2-AA) at high pH compared to a aromatic dipeptide systems where structure is clearly present pre-gelation (2-FF). Cuvette 0 (blue) and 90 (green) degree orientations are shown indicating contributions of orientation effects from self-assembly alignment.

8.2.2 Post-gelation general observations

Figure 8.4 shows the gels formed after overnight incubation post GdL (8.9 $mgmL^{-1}$) addition. The systems' ability to form self-supporting gels was examined by vial inversion: 2-AA, Br-AA, Br-AV, CN-AV, 2-FF, Br-FF, CN-FF and CN-FV form self-supporting clear gels. 2-FV and Br-FV form self-supporting turbid gels, CN-AA forms a turbid suspension that fails the vial inversion test and 2-AV is an exceptional case where upon pH decrease no change in viscosity is initially observed but after extended incubation, gelation proceeds and then later disintegrates. The turbid gels are here classified as those that are opaque and less likely to pass the vial inversion test but may be further characterised by cross-polarised light microscopy (see Appendix III-iii). From this analysis, turbid gels can be described as falling into two categories, either appearing as precipitate (2-FV) or crystalline material (CN-AA). Presumably these systems are unable to maintain a clear self-supporting gel morphology due to the presence of these phase separations. Some systems that appear to make clear self-supporting hydrogels are delicate and, in the cases of Br-AA and Br-AV, are easily disrupted upon manual manipulation indicating a low tolerance to strain (see Appendix III-iii). Ultimately, it is the aim of this work to connect the structure of these systems in the gel phase to these observations and their associated material properties, but also give insight into what underlies the differences in material formation as discussed in the following sections.

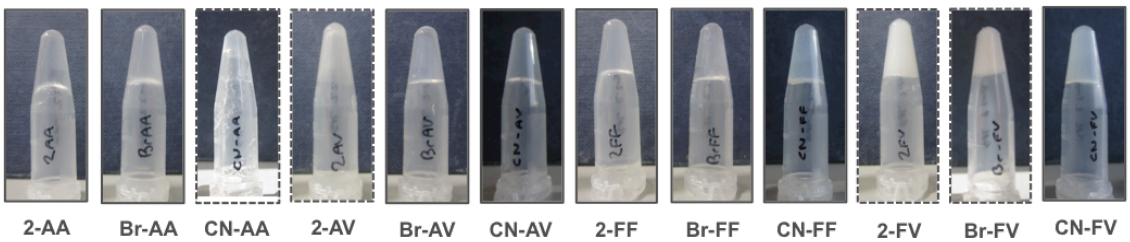


Figure 8.4: Photographs of the hydrogels formed by the dipeptide library after overnight incubation with 8.9 $mgmL^{-1}$ GdL. Gels that are maintained upon vial inversion are deemed to be self-supporting. The turbid gelators CN-AA, 2-FV and Br-FV are marked by a dashed outline. 2-AV is clear but not self-supporting.

8.2.3 Aliphatic dipeptide assembly and structure by CD

Following on from the spectroscopic and biophysical characterisation of the dipeptides at high pH (10.5), the assembly and structure of the aliphatic and aromatic dipeptides was followed by CD over the course of gelation induced by addition of GdL (8.9 $mgmL^{-1}$). The CD exhibited by the aliphatic dipeptides over the course of self-assembly and gelation is shown in Figure 8.5 (full peak analysis in Appendix III-iv). The spectra do not permit simple interpretations that might be applied to a polypeptide CD spectrum however initially basic chiral analysis can be made. Systems containing the same dipeptide sequence have similar spectral shape and peak signs indicating that the chirality of the dipeptide systems are related and determined by the peptidic portion of the dipeptide molecule. Each system displays an increase in their respective characteristic CD signals over the course of assembly and gelation, indicating an increase in the structures they arise from. The far-UV transitions attributable to the dipeptide backbone across the whole aliphatic library occur at 195.2 nm (±5.5) and 210.5 nm (±1.9) for the positive and negative peaks respectively. These have previously been attributed as indicating ß-sheet formation (see Chapter 7) and this is supported by Fourier transform infrared spectroscopy (FTIR) data indicating so (see Appendix III-v). Evidently, the structure of the peptidic portion of the aliphatic dipeptides are closely related for these peaks to be similar. The variation that is present may be due to variability in the precise environment that the dipeptide is in or perturbation of the electronic excitability of the amide bonds due to naphthalene conjugation.

The systems CN-AA and 2-AV were found to reproducibly have a very low signal-to-noise ratio or perhaps no CD signal. This phenomena will require further investigation, as the dipeptide is clearly assembling as judged by the corresponding $HT[V]_{sub}$ traces.

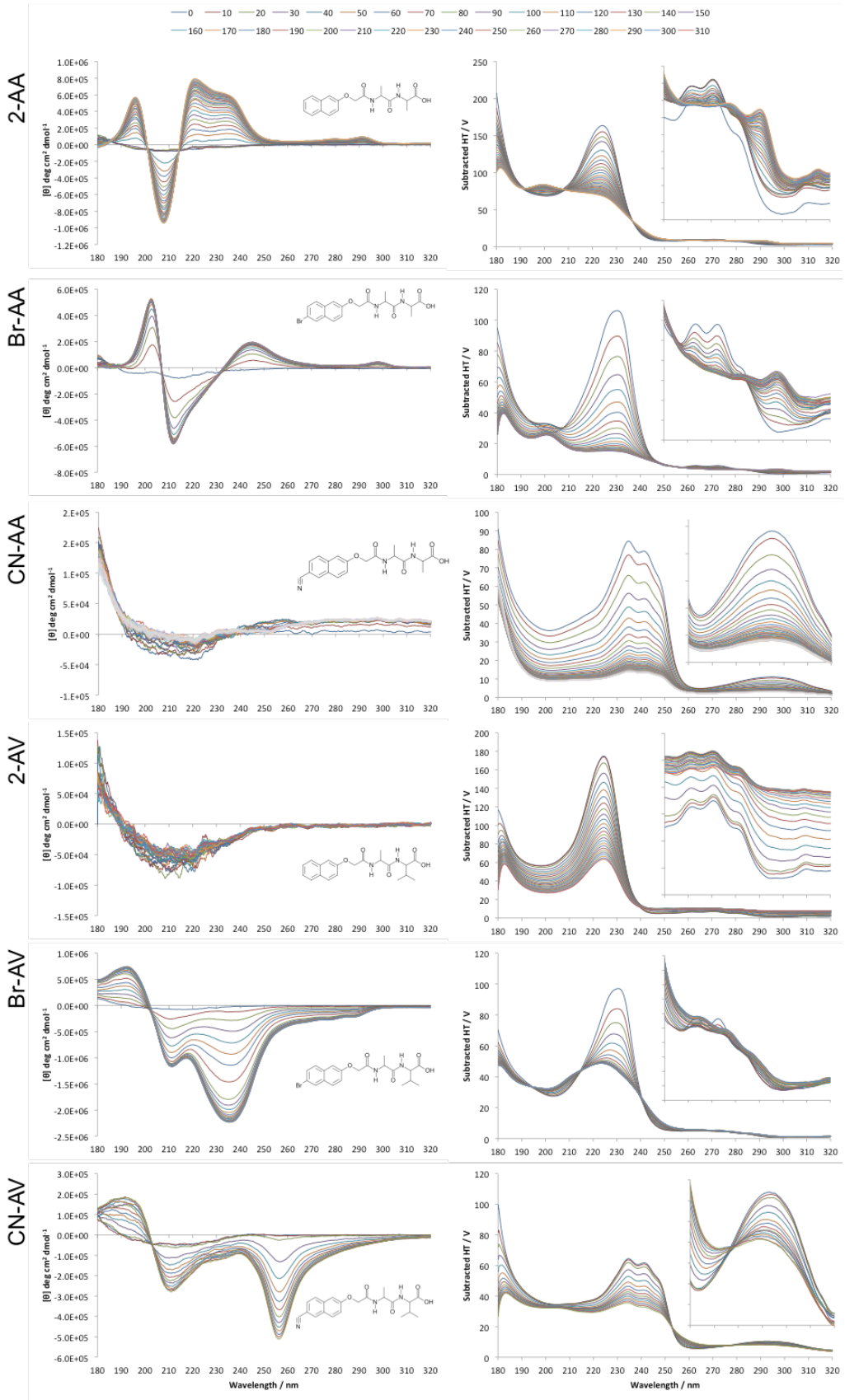


Figure 8.5: Time course CD spectra collected over the course of gelation of the aliphatic dipeptides. The legend shows the time after GdL addition in minutes. No LD artefacts were detected in the spectra by cuvette rotation. For tabulated CD signals see Appendix III-iv.

In the near-UV, the presence of CD transitions due to the naphthyl groups (>~220 nm) are observed. This indicates a close association of these groups into π-π stacking arrangements over the course of self-assembly. The formation of π-π stacking is evidently unaffected by the 6-position naphthyl substitution. The aromatic stacking is supported by the concurrent hypochromic shifting of the B_b transitions observed in the absorbance profiles derived from the $HT[V]_{sub}$; this phenomena also being observed in the systems which are CD negative. In this region the CD bands clearly shift in accordance with the absorbance profile derived from the $HT[V]_{sub}$ further indicating the origin of these CD bands. Addressed in latter sections, the excitonic natures of these transitions with regard to spectral shape appear to be governed by the dipeptide sequence.

8.2.4 Aromatic dipeptide assembly and structure by CD

The same investigation of the local packing and chirality was performed for the aromatic dipeptides using CD. Figure 8.6 show representative spectra for the assembly and gelation of the aromatic dipeptides (full peak analysis in Appendix III-iv), the relative intensities of the CD bands were found to sometimes vary between experiments. The majority of spectra for the gelation of the aromatic dipeptides do not exhibit the same development of signals as the aliphatic dipeptide systems. This is attributable to the presence of self-assembled material prior to gelation, consistent with the CD of aromatic dipeptides at high pH (see Figure 8.3). This again indicates that self-assembly is distinct from gelation. All spectra, except for CN-FF, were found not to be the subject of LD artefacts as might be expected for fibrillar systems. Some spectra are of assembly dipeptide with Thioflavin-T (Br-FF, 2-FF and 2-FV) but this was confirmed to not affect the observed signals.

Similar trends in spectral shapes are not as easily found between the systems, except between 2-FF and Br-FF. It may be tempting, as with the aliphatic dipeptides, to suggest that the chirality reported by the CD is determined by the dipeptide sequence but CN-FF shows a different

spectra to its un-substituted and bromonaphthalene conjugated homologues. No correlation is evident in the CD spectra between FV containing dipeptides. It is surprising and interesting that a related dipeptide 2-FG bears relation to 2-FV (see Appendix III-vi). This indicates that the same local assemblies are formed in 2-FG as they are in 2-FV. However, the latter is an efficient gelator whereas the former is not, further indicating that gelation propensity is distinct to assembly propensity and two comparably structured fibres can form different material phases.

Analysis of the CD may be complicated because of overlapping phenylalanine transitions and complex exciton couplets but by the end of assembly and gelation each has a high-energy positive CD band in the π-π^* region expected for amide bonds at 193.9 nm (± 2.8) comparable to the same band observed for the aliphatic dipeptides. The negative low energy component expected for ß-sheets may be overlapping with the expected 210 nm L_a phenylalanine transition but this is resolved in the systems 2-FF, Br-FF and CN-FV at 207.7 nm (±2.2) comparable to the aliphatic dipeptides. It appears as though the conformation of the aromatic dipeptide backbones are comparable to those of the aliphatic dipeptides in configuration and ß-sheet formation. This is also corroborated by FTIR data indicating ß-sheet formation (see Appendix III-v).

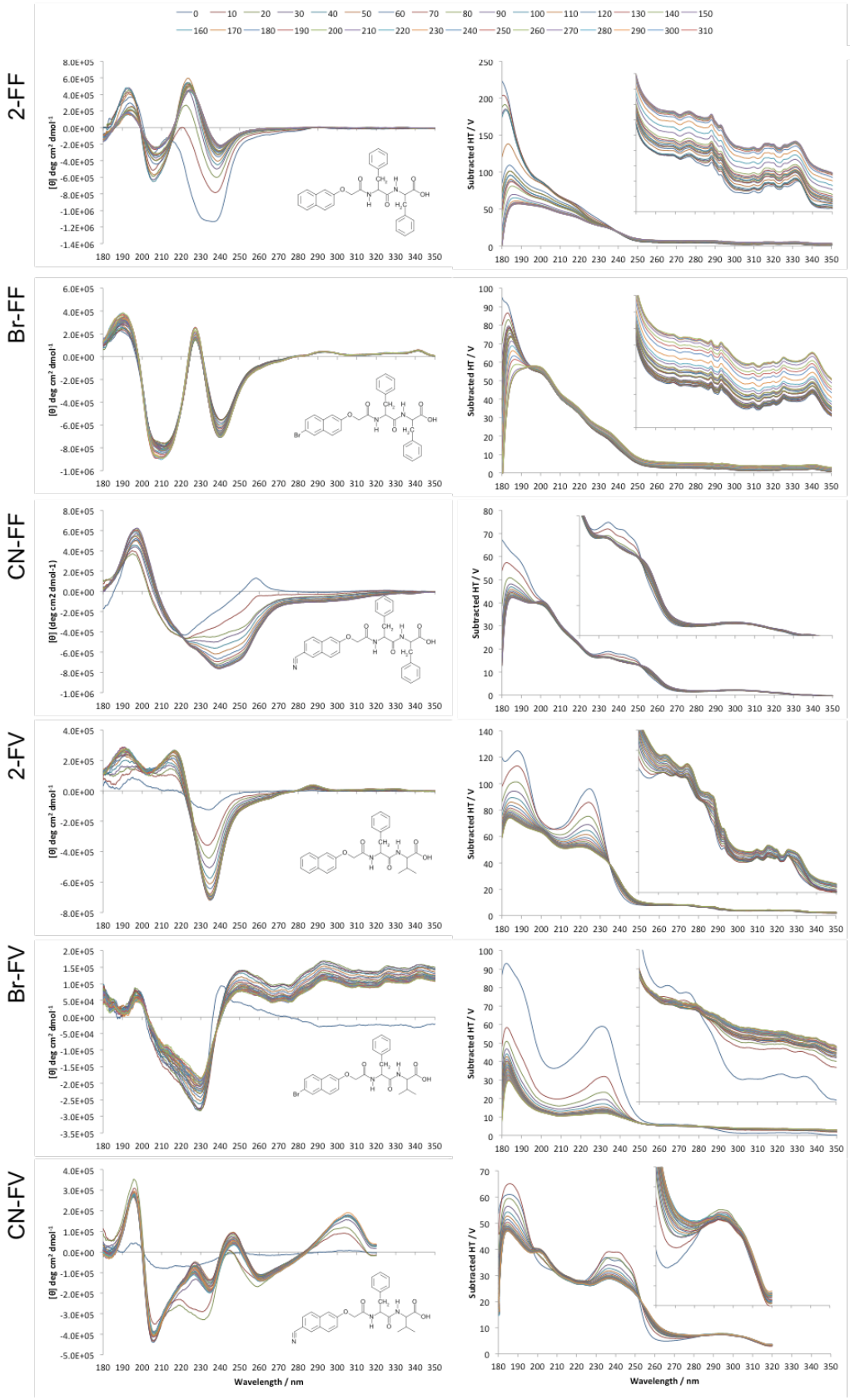


Figure 8.6: CD spectra collected over the course of assembly and gelation for aromatic dipeptides. The legend shows the time after GdL addition in minutes. Br-FF, 2-FF and 2-FV spectra are of dipeptide with ThT. No LD artefacts were detected in the spectra by cuvette rotation except for CN-FF (see Appendix III-vii). For tabulated CD signals see Appendix III-iv.

8.2.5 Spectroscopic structural interpretations

Exciton coupling effects

The assessment already made of the CD based on general spectral shape, amide transition peak position and the observation of naphthyl chiral stacking is useful in so much as what has previously been described for these classes of systems (see Chapter 7). A more thorough analysis may consider the precise origins of the exciton coupling, in particular involving the naphthalene groups as successfully applied to Br-AV previously (see Chapter 7) and as described in Chapter 3.

Helicity Analysis

The helicity of coupled chromophores is established by analysis of the Cotton effects, that is the rapid change in the sign of the CD signal, present in the spectra (see Section 3.2.6). Comparing the cases of Br-AV and Br-AA, it is immediately apparent that the negative Cotton effect around the B_b (232 nm) naphthyl transition of Br-AV is observed but a positive Cotton effect is observed for the same B_b (232 nm) naphthyl transition of Br-AA, indicating a left and right-handed helical arrangement for each dipeptide system respectively (Rodger and Norden, 1997). However, in line with this, similar exciton bands might be expected for the naphthyl groups in CN-AV and 2-AA though they are not easily identifiable. A single negative band is observed at 256.4 nm with no observable positive couplet for CN-AV whilst two positive components are observed at 221.1 and 235.0 nm for 2-AA. Thus, these phenomena cannot currently be rationalised within the same theoretical framework as that applied to Br-AA and Br-AV.

Assignment of the exciton bands from the naphthyl groups for the aromatic dipeptides can be done so with more confidence due to the available wavelength data, the $HT[V]_{sub}$ and UV/Vis data for the stocks granting knowledge of the B_b naphthalene long axis transition wavelength. In each case the naphthyl B_b transition is observed to split into a low and high-energy components as shown in Table 8.5. A comparison of the expected midpoint of these bands to the B_b transition identified by UV/Vis analysis of the aromatic dipeptide stocks confirms that

these transitions are exciton couplets of the long axis transition of naphthalene. The exciton splitting phenomenon in the CD is, as with the aliphatic dipeptides, consistent with the close interaction of these groups upon self-assembly. Further, the signs of the exciton components are indicative of a helical chiral arrangement of the naphthyl groups in the fibrillar assemblies; a left-handed helicity is apparent for 2-FF, 2-FV and Br-FF whilst a right-handed helicity is observed for Br-FV and CN-FV. Peaks are not easily assignable for CN-FF due to spectral shape and the presence of LD artefacts.

Where exciton coupling in the CD gives an indication of the chirality and helical supramolecular arrangements of these molecules in the fibrillar phase, the exciton coupling exhibited by the absorbance traces derived from the $HT[V]_{sub}$ is useful in differentiating these systems from one another.

System	Time / mins	Low-energy / nm	Sign	High-energy / nm	Sign	Expected zero / nm*	$HT[V]_{sub}$ / nm	Abs / nm	Assignment	Helicity
Br-AA	10								Nap Bb	
	120	245.0	+	212.0	-	228.5	228.9	232.0	Nap Bb	R
	270	245.5	+	211.8	-	228.7	226.9	232.0	Nap Bb	R
Br-AV	10								Nap Bb	
	120	235.5	-	211.2	+	225.7	224.0	232.0	Nap Bb	L
	240	235.7	-	211.4	+	225.9	223.4	232.0	Nap Bb	L
2-FF	10	236.2	-	212.7	+	224.5		225.0	Nap Bb	L
	120	239.6	-	224.4	+	232.0		225.0	Nap Bb	L
	270	239.9	-	225.1	+	232.5		225.0	Nap Bb	L
2-FV	10	233.9	-	214.7	+	224.3	224.5	225.0	Nap Bb	L
	120	235.1	-	216.3	+	225.7	220.0	225.0	Nap Bb	L
	200	235.0	-	216.7	+	225.9	219.4	225.0	Nap Bb	L
Br-FF	10	240.2	-	227.7	+	234.0		232.0	Nap Bb	L
	120	239.9	-	227.4	+	233.7		232.0	Nap Bb	L
	230	239.7	-	227.1	+	233.4		232.0	Nap Bb	L
Br-FV	10	240.6	+	228.3	-	234.5	230.8	232.0	Nap Bb	R
	120	251.1	+	229.8	-	240.5	231.0	232.0	Nap Bb	R
	200	252.0	+	230.1	-	241.1	230.4	232.0	Nap Bb	R
CN-FV	10	244.1	+	230.7	-	237.4	235.3	236.0	Nap Bb	R
	120	246.8	+	234.6	-	240.7	236.7	236.0	Nap Bb	R
	230	246.7	+	235.1	-	240.9	236.5	236.0	Nap Bb	R

*Table 8.5: A tabulated summary of the exciton coupling from naphthalene groups of the aromatic dipeptides. * Expected zeros are calculated as midpoints. Some $HT[V]_{sub}$ values are not measurable because of hypsochromic shifting due to self-assembly. Abs values are taken from Figure 8.2. L – left-handed, R – right-handed.*

The $HT[V]_{sub}$ of the systems are almost identical to the UV/Vis of the pre-assembly stocks shown in Figure 8.2 as previously observed for Br-AV (see Chapter 7), thus the interpretations of the UV/Vis data are also applicable here and report on the electronic excitation properties of the dipeptide systems as they assemble. In each case the intense B_b absorbance band from naphthalene decreases over time, consistent with the naphthalene groups moving into close proximity of one another.

Geometry analysis

The naphthalene B_b absorbance band also blue shifts between 3.5 – 8.0 nm (2-AA, Br-AA, Br-AV and 2-FV) indicative of face-to-face aromatic stacking interactions for this axis. Other B_b blue shifts are small (<1 nm) or not apparent as for CN-AA, 2-AV and Br-FV possibly indicating a θ value of 54.7° between the long axis transitions or a face-to-face arrangement with variation away from the ideal geometry to abolish stokes shifting. As exceptions, any Stokes shifts present for 2-FF, Br-FF and CN-FF cannot be accurately measured and in the case of CN-FV red shifting of the B_b transition is observed. The hypochromic shifts noted in every case clearly indicates increasing aromatic proximity over the course of assembly but the variability in hypsochromic shifting implies subtle differences in the precise aromatic geometries in each assembled system. In a similar line of argument as made in Chapter 7, based on the exciton coupling models (Kasha, 1963, Kasha et al., 1965) for the majority of the dipeptide systems the long axes of the naphthyl groups lie between 90 – 54.7° as found for Br-AV (see Chapter 7).

Contrastingly the $HT[V]_{sub}$ spectral shapes infer a far UV absorbance band initially below 180 nm which undergoes hypochromism and red shifting to approximately 183 nm. The wavelength region that this occurs over implies that is arises from the peptide bond but what structural change it represents is uncertain. Further changes in the electronic excitability of the dipeptides are observed upon assembly in the development and loss of UV/Vis peaks with time. In addition to the general observations described, the changes to the short axis transitions of the

naphthalene are more complex. All are observed to undergo hypochromic shifting over the course of assembly though the Stokes shifting inferring details about the exciton coupling geometry of these systems does not form a clear trend amongst the library. Br-AA, CN-AA and 2-AV exhibit no short axis transition shifting indicating a θ value of 54.7° between the short axis transitions or a face-to-face arrangement with variation away from the ideal geometry to abolish Stokes shifting. Br-AV naphthyl short-axis red shifting indicates an end-to-end short axis arrangement with θ at 0 – 54.7° and CN-AV blue shifting indicating the converse face-to-face arrangement with θ at 54.7 – 90.0°. 2-AA is an unusual case where both blue and red shifting is observed in the absorption bands attributable to the short axis transitions. The same analysis applied to the aromatic dipeptides is challenging due to the presence of pre-assembled material. This is particularly true for the diphenylalanine containing dipeptides but careful comparison of the stock CD (see appendix III-i & ii) and stock UV/Vis (Figure 8.1) to the $HT[V]_{sub}$ of the FV containing dipeptides reveals some useful geometrical information. The B_b 225 nm naphthalene transition of 2-FV blue shifts to 219.4 indicating a face-to-face long axis geometry whilst the short axis naphthalene transitions of 271 and 311 nm red shift to 274.5 and 315.6 nm respectively indicating an end-to-end short axis naphthalene geometry. The Stokes shifts associated with the naphthalene transitions of Br-FV and CN-FV are either small or rapidly abolished by self-assembly and so are not interpreted here.

The aliphatic dipeptides appear to be grouped in terms of their chiral conformation by their dipeptide sequence. The aromatic dipeptides cannot be as simply categorised, though certainly 2-FF and Br-FF are related in terms of spectral character and also the helicity of the assembled naphthalene groups. 2-FV is seen to share the same helical arrangement of naphthalene groups as 2-FF and Br-FF although the spectral shape is quite different. Then Br-FV and CN-FV, although having distinct spectral characters share the same naphthalene assembled helicity whilst CN-FF is not easily characterised with any of the systems. Despite the difficulty in characterising the naphthalene geometries in these systems due to pre- or rapid assembly the low wavelength exciton couplets do seem to indicate the same peptidic structure as those found

for the aliphatic dipeptides. To understand these structural features in the context of the material formed the following sections describe the visual and structural characterisation of the dipeptide library.

8.2.6 Aliphatic and Aromatic morphology analysis by TEM

Irrespective of the material observations made on the gels, each of the dipeptides of the library were found to have a fibrillar structure in the gel phase as shown in Figure 8.7. When visually compared, AV systems appear morphologically similar to one another and certainly distinct from AA containing dipeptides. 2-AV exhibits noticeable lateral fibril association as observed for the previously characterised Br-AV (see Chapter 7). The minimum fibril width for these systems was determined to be 17.1 nm (SD ±1.9 n=4) & 17.2 nm (SD ±1.9 n=4) nm respectively. It is tempting to compare AA containing systems morphologically but only with regard to the lesser degree of lateral fibril association they show. The minimum fibril width of 2-AA is 9.3 nm (SD ± 0.9 n=10) whereas the minimum width of Br-AA fibrils is 21.7 nm (SD ± 2.4 n=9) showing parallel striations indicative of a basic fibril structure that is either two parallel protofilaments or a nanotubular arrangement (indicated by arrows). The minimum fibril width of CN-AA is 4.3 nm (SD ± 0.9 n=15) but larger helical-like fibrils on average 13.7 nm (SD ± 2.0 n=8) in diameter are also observed. CN-AV also adopts a helical-like arrangement of an average diameter of 14.8 nm (SD ± 2.2 n=8). 2-FF, Br-FF and CN-FF form thin filaments of 12.7 nm (SD ± 2.3 n=10), 9.3 nm (SD ± 0.7 n=8) and 9.5 nm (SD ± 1.1 n=8) respectively. Thicker fibrils of 28.3 nm (SD ± 3.6 n = 10) are also observed for CN-FF. 2-FV and Br-FV form thin filaments of 8.5 nm (SD ± 0.9 n=10) and 11.3 nm (SD ± 1.9 n=8) respectively whilst CN-FV filaments have a helical morphology of 22.6 nm (SD ± 2.9 n=7) at the widest point. Br-FV and CN-FV are observed to also adopt relatively large nanotubular structures of approximately 200 and 150 nm respectively. Compared to the aliphatic dipeptides, the aromatic dipeptide fibrils appear to exhibit less evidence of lateral association of fibrils although cases occur in 2-FF and 2-FV.

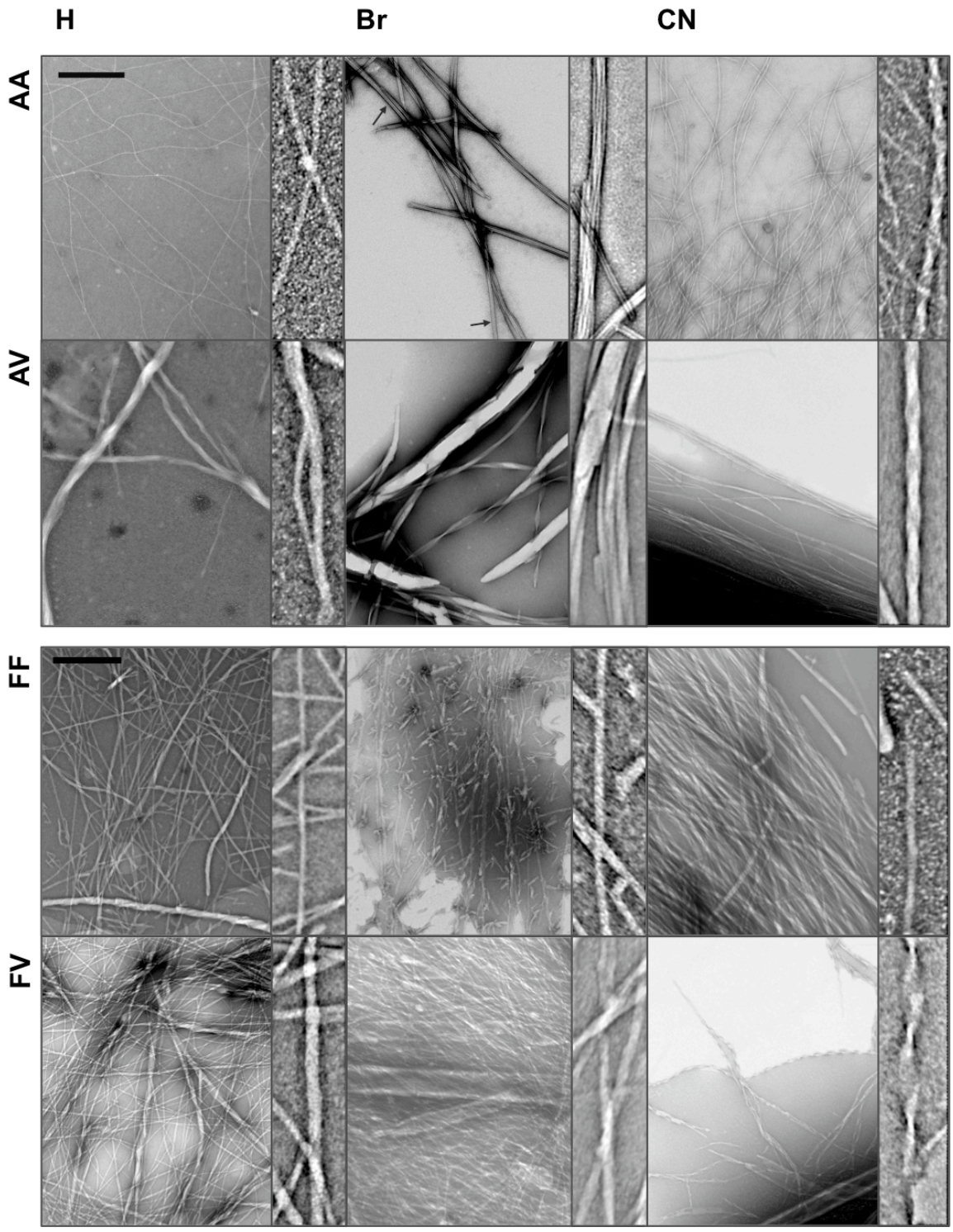


Figure 8.7: TEM of in situ prepared gels of the dipeptides. Each system is observed to be fibrillar but variations in morphology are also apparent. The scale bar represents 500 nm. The inserts show FFT crops to illustrate the most representative morphologies – not to scale.

8.2.7 Aliphatic and aromatic structural packing by XRFD

Where each aliphatic and aromatic dipeptide system forms long fibrils this enabled the alignment of these and preparation of samples for XRFD (see Appendix III-viii), the diffraction from which are shown in Figure 8.8. Fibre diffraction patterns were collected from multiple alignments and at different exposure points where considered appropriate. The data presented shows the signals for the highest quality patterns with respect to alignment, signal-to-noise ratio and information content. All fibre diffraction patterns except for that of Br-AA were collected using *in situ* alignment. Inconsistent with the majority of the dipeptides, the *in situ* sampling method was unable to efficiently sample Br-AA fibres for XRFD. It was found that sampling from a pre-formed gel produced a high quality semi-crystalline diffraction pattern. This could be consistent with sampling crystalline material from the gel matrix where presumably crystals are more soluble than large self-assembled fibrils. The case of 2-AV is also distinct whereby an alignment prepared in D_2O was found to have more directional information than its H_2O counterpart. The reflections of 2-AV in H_2O and D_2O were confirmed to overlay suggesting minimal structural differences (data not shown) and so the superior diffraction pattern of 2-AV in D_2O is used for analysis. Some systems failed to effectively align and so reflections may be radially averaged but where reasonable these have been assigned to the axis they are expected to occur on (see Appendix III-ix). It is perhaps interesting that inspection of the fibre diffraction patterns from aromatic dipeptides reveals a greater degree of radial averaging indicating a difficulty in aligning these systems. The difficulty could be due to increased entanglement of fibres which might be correlated to the material properties of the phenylalanine dipeptides but in only two out of six cases do the phenylalanine containing dipeptides have significantly different G' values to the rest of the dipeptide library (see Table 8.1).

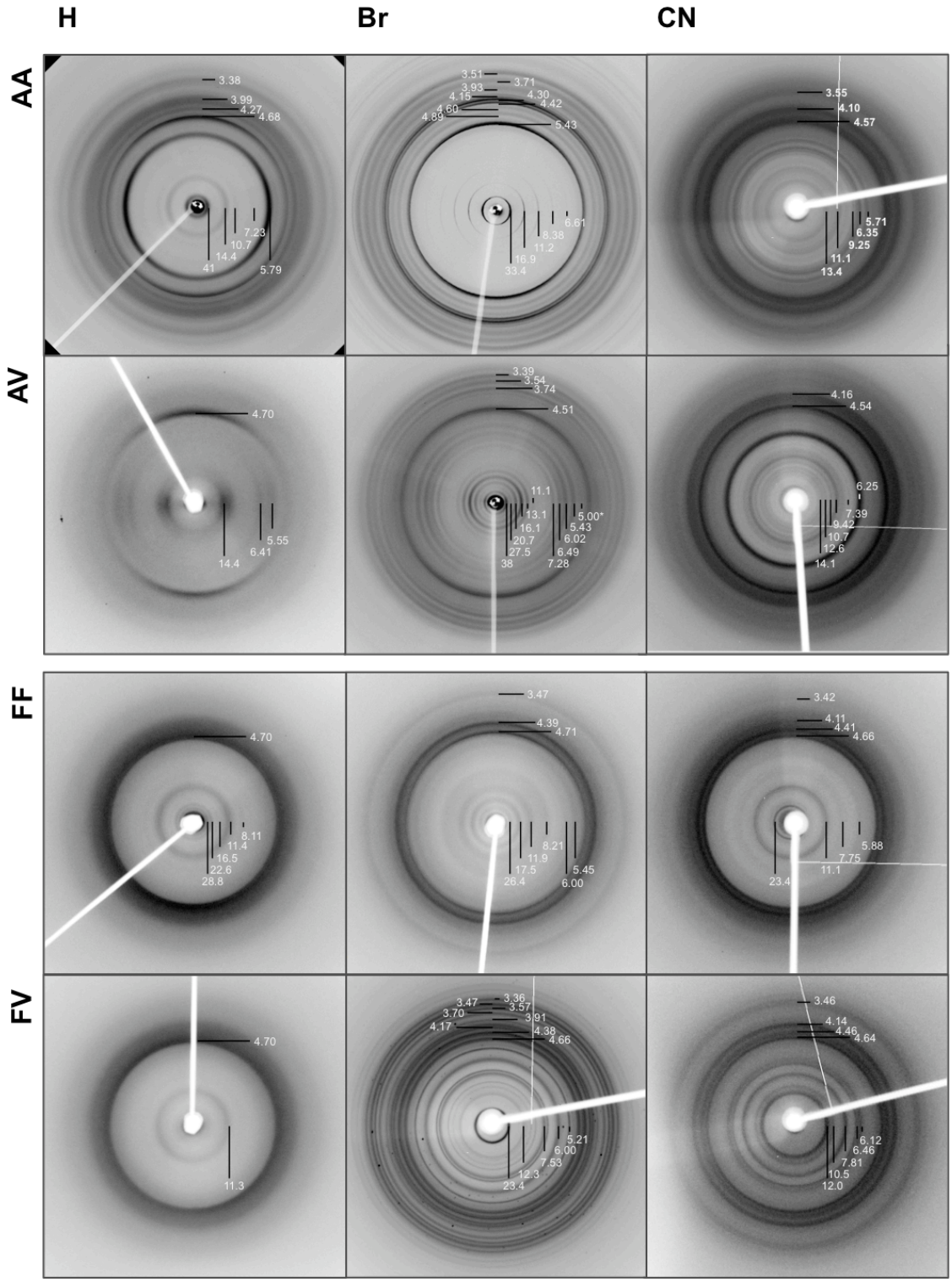


Figure 8.8: X-ray fibre diffraction exhibited from alignments of aromatic and aliphatic dipeptides. Reflections tabulated in Appendix III-ix. All fibre axes are vertical.

Indicated by the strong meridional reflections, un-substituted naphthalene conjugated aliphatic dipeptides are seen to have a repetitive spacing along the fibre axis of between 4.67 – 4.72 Å whereas bromonaphthalene and cyanonaphthalene aliphatic dipeptides have a shorter spacing of 4.28 – 4.54 Å. These distances are both compatible with a hydrogen bonding distance and are reminiscent of the meridional reflections observed for cross-ß amyloid systems. An exceptional and intense 5.43 Å meridional reflection is observed for Br-AA but this is too long for hydrogen bonding. In contrast, for the aromatic dipeptide systems the major meridional distance occurs between 4.60 – 4.73 Å across the entire aromatic library; much more consistent than the spacing of the aliphatic dipeptides which vary by 0.44 Å. Perhaps the aromatic phenylalanine groups dominate the stacking preference that these systems adopt.

A comparison of the equatorial reflections of the aliphatic systems finds large differences indicative of differing repetitive spacings perpendicular to the fibre axis on local (up to 20 Å) and long-range distances (up to 41 Å). The discrepancy in the equatorial reflections is consistent with differing lateral long-range packing arrangements, as also observed in the TEM, although similarities may be present in the local packing separations. The ß-strand stacking and naphthyl packing distance of Br-AV was found to be 6.02 Å (+0.04/-0.04) (see Chapter 7), in some cases it is clear that a similar repetitive spacing perpendicular to the fibre axis is present in other systems. For instance 2-AA exhibits an intense 5.79 Å (+0.05/-0.05) that may also be consistent with a ß-sheet separation. Approximately equivalent equatorial reflections for the aliphatic dipeptides are observed as follows 2-AV – 5.55 Å (+0.07/-0.07), Br-AV – 6.02 Å (+0.04/-0.04), CN-AA – 5.71 Å (+0.05/-0.05), CN-AV – 6.25 Å (+0.00/-0.00) and may also be consistent with a ß-sheet separation. Due to the bulky phenylalanine residue aromatic dipeptides might be expected to have a larger ß-sheet separation (Fandrich and Dobson, 2002) where in the following systems the shortest shared equatorial separation is ~11 – 12 Å as in the following systems 2-FF – 11.39 Å (+0.08/-0.08), 2-FV – 11.32 Å (+0.37/-0.35), Br-FF – 11.94 Å (+0.02/-0.02), Br-FV – 12.26 Å (+0.14/-0.14), CN-FF – 11.09 Å (+0.27/-0.26) and CN-FV – 11.98 Å

(+0.11/-0.11). These distances may be consistent with the ß-sheet separation in the aromatic dipeptides.

A complete detailed comparison of the aromatic equatorial reflections finds similar reflection positions but the relative intensities of these are variable. In a similar manner as to the stacking of dipeptides parallel to the fibre axis, perhaps the perpendicular structural separations in the aromatic dipeptides are also predetermined by the phenylalanine groups. However, though the structural spacings may indicate that the unit cells of these systems are similar on the perpendicular dimensions, the variable intensity indicates that the absolute structural configuration within these is variable. The major equatorial reflections exhibited by the aliphatic and aromatic dipeptide systems are shown in Table 8.6 (see all Appendix III-xi). This shows that the aliphatic dipeptides may be grouped in their lateral packing dimensions by the nature of naphthyl conjugation, whereas comparison may be made between the lateral packing dimensions of all of the aromatic dipeptides. Granted, additional equatorial reflections are present in some systems that are not in others but this could be due to a number of factors regarding the quality of fibre alignments obtained including crystallinity, degree of alignment and order. Notwithstanding this, based on these comparisons the evidence for the proposed structural groupings of the dipeptides based on naphthyl group or phenylalanine content is compelling. The same is applicable to the repetitive spacing running along the fibre axis; these taken together and considering the structural descriptions in Chapter 7 these data are consistent with a cross-ß like architecture with variable lateral packing architectures.

2-AA		2-AV		Br-AA		Br-AV		CN-AA		CN-AV	
Signal / Å	Intensity	Signal / Å	Intensity	Signal / Å	Intensity	Signal / Å	Intensity	Signal / Å	Intensity	Signal / Å	Intensity
40.52	206.95					65.27	252.00				
14.37	115.69	14.37	185.00	33.38	249.70	38.34	127.79			10.67	140.48
7.23	178.26	6.41	140.69			20.70	102.87	9.25	130.82	9.42	161.57
5.79	242.72	5.55	139.61	16.86	120.12	16.11	123.15	8.00	124.93	7.39	131.09
				11.18	86.49			6.35	169.87	6.25	239.86
				8.38	75.04			5.71	171.19		

2-FF		2-FV		Br-FF		Br-FV		CN-FF		CN-FV	
Signal / Å	Intensity	Signal / Å	Intensity	Signal / Å	Intensity	Signal / Å	Intensity	Signal / Å	Intensity	Signal / Å	Intensity
28.75	124.00			26.43	79.48			23.39	158.63		
22.57	123.67										
				17.49	96.68						
11.39	136.39	11.32	104.66	11.94	98.91			11.09	141.54	11.98	171.97
8.11	106.68			8.21	94.86						
						7.53	147.16	7.75	109.54	7.81	174.81
						7.12	147.94			6.46	171.44
						6.00	165.23	5.88	*	6.12	172.66
						5.21	187.59				

Table 8.6: A comparison of the most intense equatorial reflections exhibited by the aliphatic and aromatic dipeptides in the fibrillar phase.

Unit cell indexing

In an attempt to gain an understanding of the context of the local packing arrangements determined by the CD measurements, unit cells were determined by the *Unit Cell Optimisation* module of Clearer using the fibre diffraction data shown in Figure 8.8, the reflections that were used are shown in full in Appendix III-ix. The reflections were entered into the *Unit Cell Optimisation* module with a number of decimal places allowing for suitable variation in the reflection positions based on the systematic error associated with each reflection measured. The 2-FV diffraction pattern was not indexed where only one equatorial reflection was observed.

A true unit cell cannot be smaller than the largest reflection on a fibre diffraction pattern, however, a unit cell may be larger than the largest measurement of a diffraction pattern where reflections may be weak or occluded by the backstop. As such, this was factored for in the unit cell determination process. The unit cell search limit within the *Unit Cell Optimisation* module of Clearer determines the upper search limit that reflections will be indexed to. To ensure that the possible unit cell dimensional space was fully explored, the unit cell estimate was systematically adjusted through large to small orthorhombic estimates; including search sizes of *a* 70 – 20 Å, *b* 50 – 10 Å and *c* 4.71 – 4.30 (*c* was in each case determined by the major meridional reflection of each pattern). The lowest error matches were collated from the searches

(see Appendix III-xa) and the standard deviation around each dimension calculated. For dimensions perpendicular to the fibre axis, this was found to be high, however a smaller standard deviation was found around the unit cell dimensions when systems were considered grouped by their naphthalene conjugate (see Appendix III-xb) implying a relationship between these systems. Certainly it is clear that the un-substituted naphthalene dipeptides pack into a cell of *a* 37.7 ±3.2, *b* 15.0 ±0.37 and *c* 4.69 ±0.01 Å (with the exception of 2-FF where a better indexing was found at greater search limits: *a* 44.7 *b* 34.4 *c* 4.67 Å), whereas cyanonaphthalene dipeptides pack into a cell with the dimensions of *a* 45.0 ±3.44, *b* 15.5 ±3.34 and *c* 4.57 ±0.05 Å. Within the search limits used above, the variation around the bromonaphthalene dipeptide unit cells were still high. Since the largest equatorial reflection for the Br-AV system is 65.27 Å (+14.3/-9.93 Å), it was not appropriate to index a unit cell for Br-AV within the same search width as that used for un-substituted and cyanonaphthalene dipeptides and so the largest search limit was used (*a* 70 *b* 50 *c* 4.51 Å). The large cell search limit finds a unit cell of *a* 65.7, *b* 48.2 and *c* 4.51 Å. Using larger search limits for each bromonaphthalene dipeptide results in a closer correlation between unit cell determinations as shown in (see Appendix III-xd). The data shown here are consistent with bromonaphthalene dipeptides packing into a cell with the dimensions *a* 66.3 ±1.57, *b* 38.0 ±9.02 and *c* 4.55 ±0.18 Å.

Using large unit cell search limits, in this predictive manner, where the equatorial information does not always extend to the same resolution must be tentatively considered. It is a natural consequence of the mathematics of unit cell indexing that larger unit cells will have lower errors because smaller reflections may be indexed into larger cells in more ways than those confined by small cells, thus creating a bias towards determining larger unit cells. Of particular note, the unit cell of Br-AV is notably different to the unit cell indexed previously (see Chapter 7) but this only illustrates the difficulty in indexing such large unit cells. Nevertheless, it is interesting that trends can be found between the unit cell indexing. The unit cell solutions were systematically grouped by naphthalene conjugate and also sequence identity and the closest

correlation can be found between cells of dipeptides with the same naphthalene conjugation (see appendix III-xd) confirming the above implications.

Following this analysis the putative unit cell predictions for the dipeptide systems are shown in Table 8.7 with the measured reflections with the corresponding miller indices and predicted reflections.

System: 2-AA					System: 2-AV					System: 2-FF					System: 2-FV				
	a	*b*	*c*	*α=β=γ*		*a*	*b*	*c*	*α=β=γ*		*a*	*b*	*c*	*α=β=γ*		*a*	*b*	*c*	*α=β=γ*
Predicted unit cell:	39.91	14.72	4.68	90.00	Predicted unit cell:	35.39	15.24	4.70	90.00	Predicted unit cell:	44.66	34.44	4.67	90.00	Predicted unit cell:	N/A	N/A	4.70	90.00
	Miller Indices					Miller Indices					Miller Indices					Miller Indices			
Empirical	h	k	l	Calculated	Empirical	h	k	l	Calculated	Empirical	h	k	l	Calculated	Empirical	h	k	l	Calculated
40	1	0	0	39.91	14	1	1	0	14.00	28	1	1	0	27.27					
14	1	1	0	13.81	6.4	3	2	0	6.40	22	2	0	0	22.33					
10	4	0	0	9.98	5.5	6	1	0	5.50	16	1	2	0	16.07					
7.23	1	2	0	7.24						11.3	3	2	0	11.26					
5.7	7	0	0	5.70						8	4	3	0	8.00					

System: Br-AA					System: Br-AV					System: Br-FF					System: Br-FV				
	a	*b*	*c*	*α=β=γ*		*a*	*b*	*c*	*α=β=γ*		*a*	*b*	*c*	*α=β=γ*		*a*	*b*	*c*	*α=β=γ*
Predicted unit cell:	65.05	33.00	4.30	90.00	Predicted unit cell:	65.66	48.20	4.51	90.00	Predicted unit cell:	66.00	28.34	4.71	90.00	Predicted unit cell:	68.62	42.64	4.66	90.00
	Miller Indices					Miller Indices					Miller Indices					Miller Indices			
Empirical	h	k	l	Calculated	Empirical	h	k	l	Calculated	Empirical	h	k	l	Calculated	Empirical	h	k	l	Calculated
33	0	1	0	33.00	65	1	0	0	65.66	26	1	1	0	26.04	23	3	0	0	22.87
16	1	2	0	15.99	38	1	1	0	38.85	17.49	3	1	0	17.38	12	3	3	0	12.07
11	0	3	0	11.00	27	2	1	0	27.13	11.9	3	2	0	11.91	10.1	6	2	0	10.08
8.4	5	3	0	8.40	20	3	1	0	19.93	8.2	4	3	0	8.20	8.9	7	2	0	8.91
6.6	0	5	0	6.60	16	0	3	0	16.07	6	11	0	0	6.00	7.5	9	1	0	7.51
					13	3	3	0	12.95	5.5	12	0	0	5.50	7.1	0	6	0	7.11
					11	6	0	0	10.94						6	2	7	0	6.00
					7.3	9	0	0	7.30						5.2	9	6	0	5.20
					6.5	10	1	0	6.51										
					6	1	8	0	6.00										
					5.4	9	6	0	5.40										

System: CN-AA					System: CN-AV					System: CN-FF					System: CN-FV				
	a	*b*	*c*	*α=β=γ*		*a*	*b*	*c*	*α=β=γ*		*a*	*b*	*c*	*α=β=γ*		*a*	*b*	*c*	*α=β=γ*
Predicted unit cell:	39.89	20.25	4.57	90.00	Predicted unit cell:	46.93	14.78	4.54	90.00	Predicted unit cell:	46.84	12.49	4.54	90.00	Predicted unit cell:	46.49	14.32	4.64	90.00
	Miller Indices					Miller Indices					Miller Indices					Miller Indices			
Empirical	h	k	l	Calculated	Empirical	h	k	l	Calculated	Empirical	h	k	l	Calculated	Empirical	h	k	l	Calculated
13	3	0	0	13.30	14	1	1	0	14.10	23	2	0	0	23.42	12	2	1	0	12.19
11.1	3	1	0	11.12	12.6	2	1	0	12.51	11	2	1	0	11.02	10.5	3	1	0	10.52
9	2	2	0	9.03	10.7	3	1	0	10.74	7.8	6	0	0	7.81	7.8	5	1	0	7.80
8	5	0	0	7.98	9.4	5	0	0	9.39	5.9	7	1	0	5.90	6.5	3	2	0	6.50
6.4	2	3	0	6.39	7.4	0	2	0	7.39						6.1	4	2	0	6.09
5.70	7	0	0	5.70	6.25	4	2	0	6.25										

Table 8.7: Putative unit cell determinations for the dipeptide library. The systems are grouped by the nature of the naphthalene conjugate and include the empirically measured reflections, the indexing to their predicted cell and the calculated reflections that arise from these respectively. A unit cell is not calculated for 2-FV where only one equatorial reflection is observed.

8.2.8 Structural modelling

Fibre diffraction modelling

It is noticeable that, in a similar manner to the reports in Chapter 7, the unit cells are relatively large considering the size of the molecules that occupy them. Calculation finds that the predicted unit cells contain between 6 – 39 asymmetric units as shown in Table 8.8. Considering the geometrical constraints from the biophysical data it may be possible to propose asymmetrical unit models for the dipeptides of the library but ultimately due to the large degrees of freedom around potential molecular packing architectures the full models cannot be fully validated, making modelling the structure in the fibrillar phase problematic.

Clearly, determining the structure of the dipeptides in the fibrillar phase represents a huge challenge and thus, alternative approaches attempted to determine the structures of the dipeptides are discussed in the following sections.

Molecule	Predicted cell volume / Å3	Molecule volume / Å^3	Asymmetric units per cell
2-AA	2749	310	9
2-AV	2534	344	7
2-FF	7183	454	16
2-FV	N/A	415	N/A
Br-AA	9231	328	28
Br-AV	14273	362	39
Br-FF	8811	472	19
Br-FV	13634	433	31
CN-AA	3692	327	11
CN-AV	3150	361	9
CN-FF	2655	471	6
CN-FV	3088	432	7

Table 8.8: The numbers of molecules estimated to occupy the predicted unit cells shown in Table 8.7. Molecular volumes were calculated using the Molinspiration (Molinspiration).

8.2.9 Dipeptide crystal structures

All of the fibre diffraction data shown in this chapter are taken from alignments of fibres prepared *in situ* (with the exception of Br-AA); however some systems are uniquely observed to crystallise from the gel phase. The occurrence of crystal formation in the same solvent in which the gel is formed is a rare phenomenon, where often to crystallise a gelating molecule disparate solvents are used to promote crystallisation (Adams et al., 2010a, Marshall and Serpell, 2010). It is uncertain whether crystals grown from the fibre phase represent an independent assembly event or a phase reorganisation of the gel fibres as also questioned for other self-assembling systems (Marshall et al., 2010, Marshall and Serpell, 2010). Despite the uncertainty over the comparability of these phases, a crystal engineering approach to LMWG design and apparent structure determination is commonly adopted (Dastidar, 2008). Though to be treated with caution, crystals grown from the gel phase would be expected to better represent the native gel structure than crystals of a gelator grown in a different solvent necessary for crystallisation and thus these crystal structures may be relevant to LMWG fibre structure. The following sections describe the structures of crystal structures grown from the gel phase of the naphthalene-conjugated dipeptides 2-AA and 2-AV. 2-AA is found to crystallise from the gel phase after extended incubation. 2-AV typically enters the crystalline phase immediately. The crystallisation of these systems is reported in the manuscript in preparation by Houton *et al.* (Houton et al., 2012). Additionally, the structures of two 1-position naphthalene-conjugated dipeptides 1-GA and 1-AG not included in the main library are also described. 1-GA is found to crystallise from the gel phase after extended incubation, whereas 1-AG enters the crystalline phase immediately (Adams et al., 2010a).

2-AA crystal structure

Preliminary crystal structures of 2-AA presented an opportunity to analyse whether these represented the structure in the fibre phase. Although a fibrillar structure is not available for 2-AA, the fibre diffraction signals indicate the repetitive spacings within the fibre phase that can be compared to the lattice spacings of the crystal structure. 2-AA crystallises in an

orthorhombic unit cell of the dimensions *a* 5.79, *b* 8.61, *c* 36.39 Å in a P $2_12_12_1$ space group. From the fibre diffraction data the major interatomic distances parallel and perpendicular to the fibre axis are known to be 4.68 Å and 5.79 and 41 Å, and it is thus tempting to consider the *b* axis of the crystal structure as representing the fibre axis with comparable lateral packing spacings in the crystal and fibre phase. Indeed, this is also consistent with the configuration of the cell as shown in Figure 8.9. The peptidic portion of the molecule has the dihedral angles of Φ -62.9°, Ψ 142.6° (blue diamond) and Φ -88.9, Ψ 150.7 (red circle), for the first and second amide bonds respectively, consistent with ß-strand conformation (a). The dipeptides pack such that ß-strands are in an in-register antiparallel arrangement stacked along the *b* axis stabilised by interamide-carboxyl hydrogen bonding (b). Two aromatic interfaces are present viewed down the *a* axis including paired and displaced paired naphthalene groups (b). On the perpendicular *a* dimension ß-strands are aligned allowing carbonyl-water-carbonyl hydrogen bonding between adjacent ß-strands. Naphthalene groups are displaced by *a/2* 2.89 Å such that a direct aromatic interaction along the *b* axis (fibre axis) is not present. Instead potential aromatic interactions occur either along the *a* dimension spaced at 5.79 Å or along the *ab* plane at 4.99 Å, in each case the electron overlap is small (see appendix III-xi). The lattice is stabilised by a hydrogen-bonding network parallel to the *b* axis between the backbone amide and carboxyl groups of stacked dipeptides, with approximate inter-ß-strand spacing of 4.37 Å (b). An additional hydrogen-bonding network exists perpendicular to the *b* dimension between backbone carbonyl oxygens and molecular water stabilising adjacent ß-sheets (c).

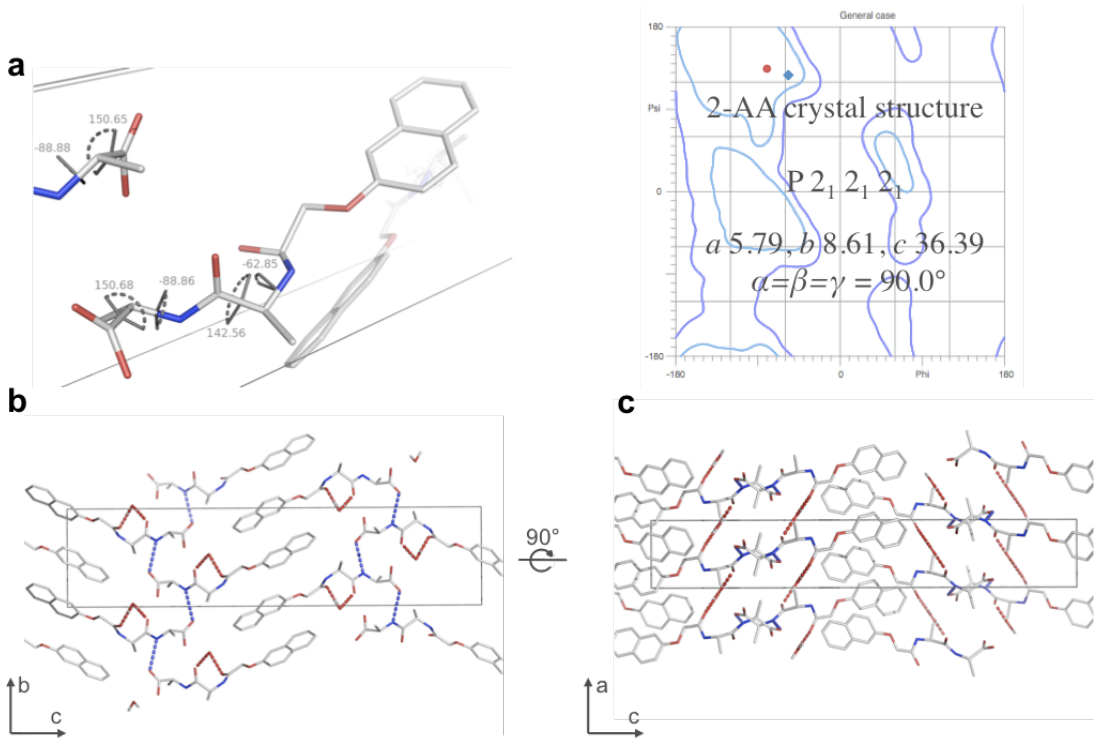


Figure 8.9: The structure of 2-AA in the crystalline phase. The peptide dihedrals classify this molecule as ß-strand (a). Viewed down the a axis, the interstrand distance is approximately b/2 = 4.31 Å stabilized by interamide-carboxyl hydrogen bonding shown in blue (b). Viewed down the b axis the distance between naphthalene pairs is 5.79 Å. ß-strands are stabilised by carbonyl-water-carbonyl hydrogen bonding shown in red (c). Coordinates provided by Dave Adams (Liverpool) and solved by Hiroyasu Sato & Akihito Yamano (Rigaku). Graphics generated in PyMol (DeLano, 2002).

To investigate whether the lattice spacing and structure it contains accurately represents the fibre phase, simulated fibre diffraction patterns were calculated from the crystal structure. Simulation assuming the *b* axis represents the fibre axis is shown in Figure 8.10. It was found that overall the crystal lattice packing and structure contained within did not fully reproduce the reflections found in the experimentally collected fibre diffraction pattern and so ultimately the crystal structure does not represent the fibre structure, they may, however, be subtly related. The simulated pattern does contain some reflections that in their spacing (Å) are also found experimentally but are discrepant in the axes that they occur on. This implies that some repetitive structural separations present in the gel phase are also present within the crystal structure but may be aligned to different crystallographic axes.

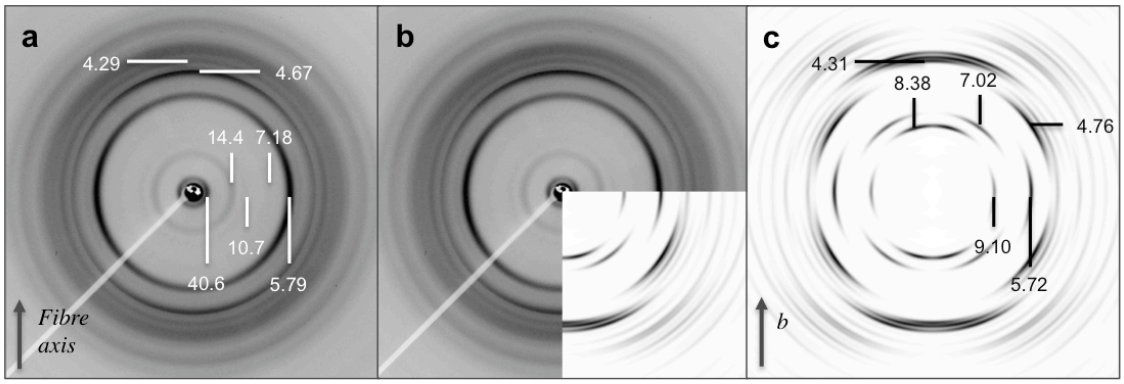


Figure 8.10: Experimental fibre diffraction collected from an in situ prepared alignment of fibres of 2-AA (a), compared to a quadrant of the simulated fibre diffraction pattern of the crystal structure of 2-AA (b). Inspection of the whole simulated pattern reveals the occurrence of some reflections that are comparable in their position but are aligned to different axes than in the experimental pattern (c). The meridian corresponding to the fibre axis is vertical, measurements are made in Ångstroms.

Close inspection of the patterns finds that the relative intensity of the simulated reflections arising from long-range spacings (>20 Å) are low such that they are not observed. The meridional and equatorial reflections are fundamentally different and many additional reflections occur off-axis indicative of the complex alignment of these structural features within the unit cell. The 5.79 Å equatorial may be reproduced as a 5.72 Å simulated reflection, corresponding to the di-peptide and naphthyl packing distance. The principle meridional reflection occurs at a much shorter distance of 4.31 Å suggesting the crystal structure is more compact along the fibre axis than in the fibre phase. It may be significant that an off-axis 4.76 Å (indexing [111]) reflection that is comparable to the meridional 4.67 Å reflection of the experimental fibre diffraction is observed. Figure 8.11 shows that the 4.76 Å reflection arises from the unit cell dimensions (a, ci and di) but that the structure strongly modulates the intensity of this reflection (b, cii and dii) indicating its presence in the crystal structure. As yet this has not been assigned to a repetitive interatomic separation within the crystal structure but the indication of an off-axis separation comparable to the interatomic separation running along the fibres in the fibre phase is no less interesting.

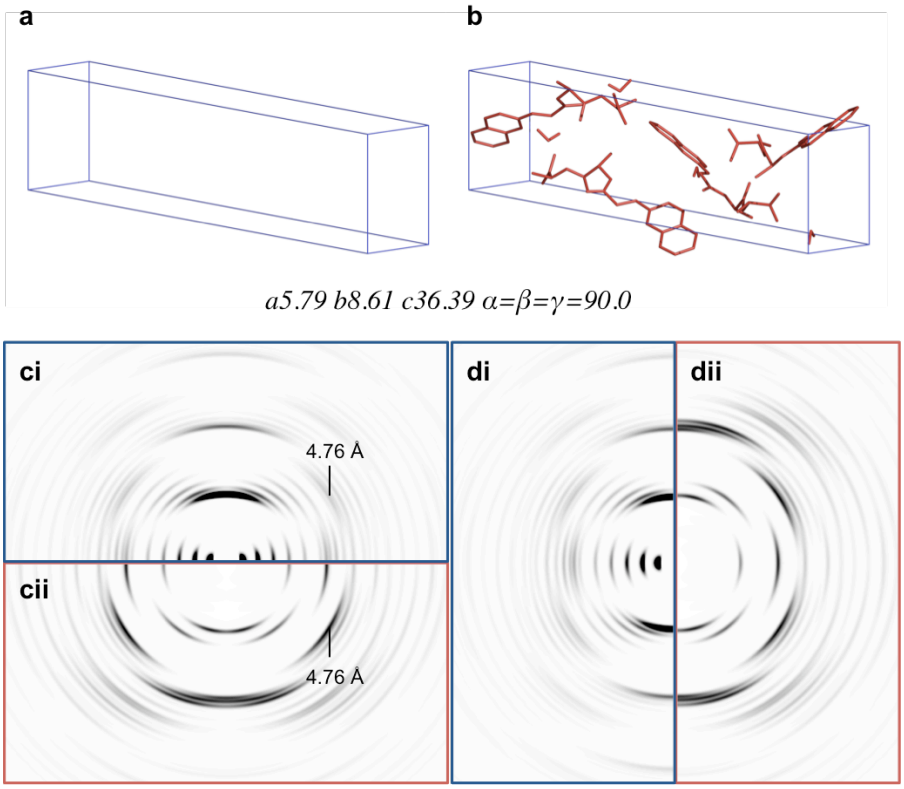


Figure 8.11: The unit cell of the 2-AA crystal structure shown in blue (a) is occupied by four asymmetric units shown in red (b), in a fibre texture the cell would give rise to the diffraction reflections shown in blue (ci and di) but the structure within this modulates the intensities of these reflections to give rise to the simulated pattern shown in red (cii and dii). The modulation of the 4.76 Å off-axis reflection implies this is a structural spacing present in the crystal structure but aligned on a complex axis. To become meridional as in the experimental fibre diffraction this may realign itself uniaxially.

2-AV crystal structure

In a similar manner to 2-AA, crystals of 2-AV were prepared. Figure 8.12 shows the packing geometry of this molecule in the crystalline phase. The asymmetric unit is related to its symmetry mates (the molecules of the unit cell generated from the asymmetric unit through space group symmetry operations, see Section 3.5.2) in a C 1 2_1 1 monoclinic unit cell of the dimensions *a* 24.62, *b* 6.98, *c* 12.55, ß = 101.2°. The peptidic portion of the molecule has dihedral angles of Φ -165.5, Ψ 154.0 (blue diamond) and Φ -139.1, Ψ 168.3 (red circle), for the first and second amide bonds respectively, consistent with ß-strand strand conformation (a). Viewed down the *b* axis the ß-strands are found to interact through two different hydrogen bonding interfaces, with a separation of 7.87 Å and 4.29 Å at the wet and dry interface respectively (see Appendix III-xii).

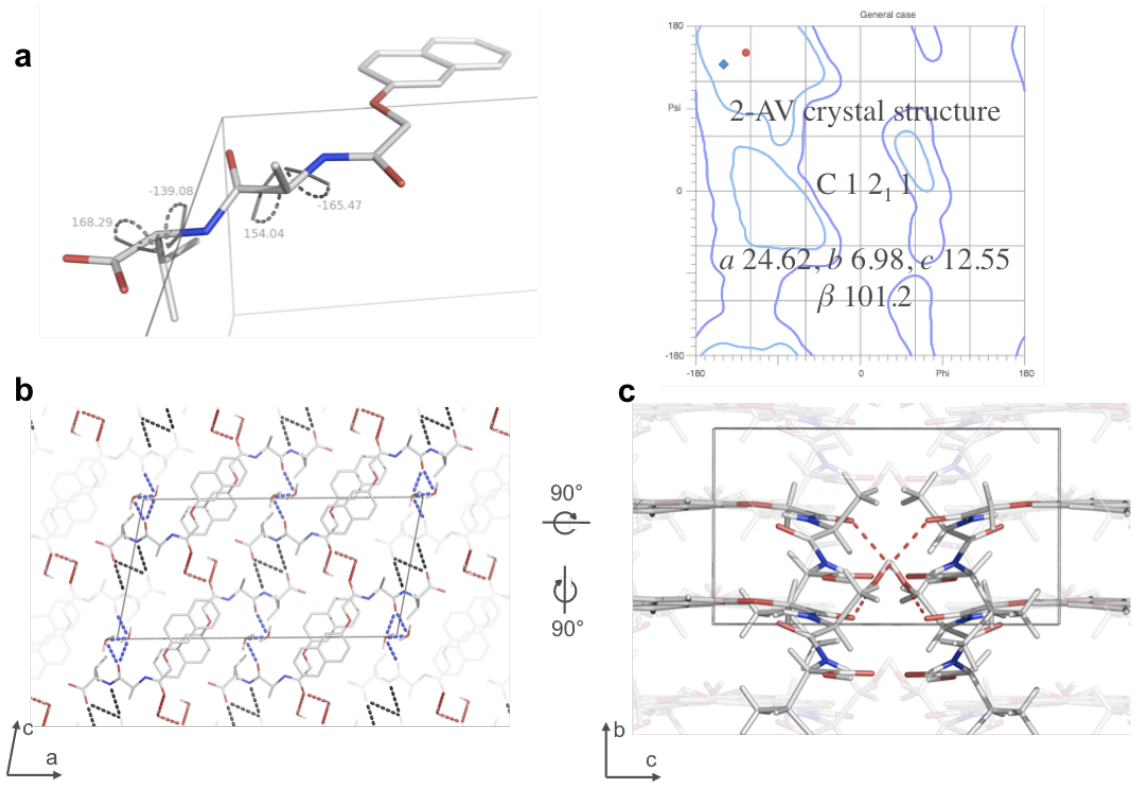


Figure 8.12: The structure of 2-AV in the crystalline phase. The peptide conformation is consistent with ß-strand (a). The lattice structure is stabilised by a complex hydrogen-bonding network. Viewed down the b axis antiparallel ß-strands are stabilised by interamide-carboxyl (grey dashed lines) and carbonyl-water-water-carbonyl hydrogen bonding (blue dashed lines) (b). A further hydrogen bonding network between adjacent ß-strands can be seen viewed down the a axis in carbonyl-water-carbonyl interactions (red dashed lines) (c). Coordinates provided by Dr. D. J. Adams (University of Liverpool) and solved by Dr. M. Schmidtmann (University of Liverpool). Graphics generated in PyMol (DeLano, 2002).

The packing produces antiparallel ß-strands that stack along the *c* axis, such that naphthalene groups are grouped at the periphery of the ß-strand interactions, however planar with the *ac* axis they are separated by *b/2* 3.49 Å with centroid-to-centroid distances of 4.71 Å and thus are unusually stacked (see Appendix III-xii). A one-dimensional hydrogen-bonding network of backbone-backbone and backbone-water-water-backbone interactions creates one-dimensional sheets (*ac* plane) of dipeptide molecules (b). Backbone-backbone hydrogen bonding occurs between carboxyl terminus oxygen and the second amide nitrogen at the dry interface and backbone-water-water-backbone hydrogen bonding occurs from the second carbonyl oxygen at the wet interface (b). An additional hydrogen-bonding network extends in the second dimension

to create the crystal lattice, with backbone-water-backbone interactions occurring from the first carbonyl oxygen extending along the *b* dimension (c).

This crystal structure assembles into a highly complicated lattice structure not as simply related to the interatomic spacings in the fibre phase. The fibre diffraction indicates spacings along the fibre axis of 4.70 Å (+0.02 / -0.02) and perpendicular to this; spacings related to 14.4, 6.41 and 5.55 Å. Close inspection of the fibre diffraction pattern from 2-AV (see Figure 8.8) reveals that the 14.4 equatorial reflection may consist of an overlapping 12.0 Å (+0.01 / -0.01) reflection comparable to the *c* dimension of the crystal structure.

Although comparisons between the dimensions of the lattice of the 2-AV crystal structure may again, like 2-AA, be compared to the interatomic separations indicated by fibre diffraction data the relationship is complex; in particular, assigning which dimension in the crystal structure represents the fibre axis not so straightforward. To further investigate this, simulated fibre diffraction patterns were calculated from the crystal structure assuming each crystallographic axis could represent the fibre axis as shown in Figure 8.13. Overall, the simulated patterns do not reproduce the experimental fibre diffraction pattern and further it is not clear which axis of the crystal structure represents the fibre axis and so comparison between the simulations and experimental data is problematic. In each case, similarities can be found in reflection positions but, as with 2-AA, the axes on which they occur are not consistent with the experimental data. For instance, the experimental meridional reflection of 4.70 Å (+0.02 / -0.02) is not reproduced as a meridional reflection, though the [012] & [311] reflections are observed off-axis at 4.62 and 4.65 Å respectively in patterns assuming *a* or *c* represent the fibre axis. In the same manner as for 2-AA,the crystal structure may be related to the fibre structure phase but is ultimately not representative.

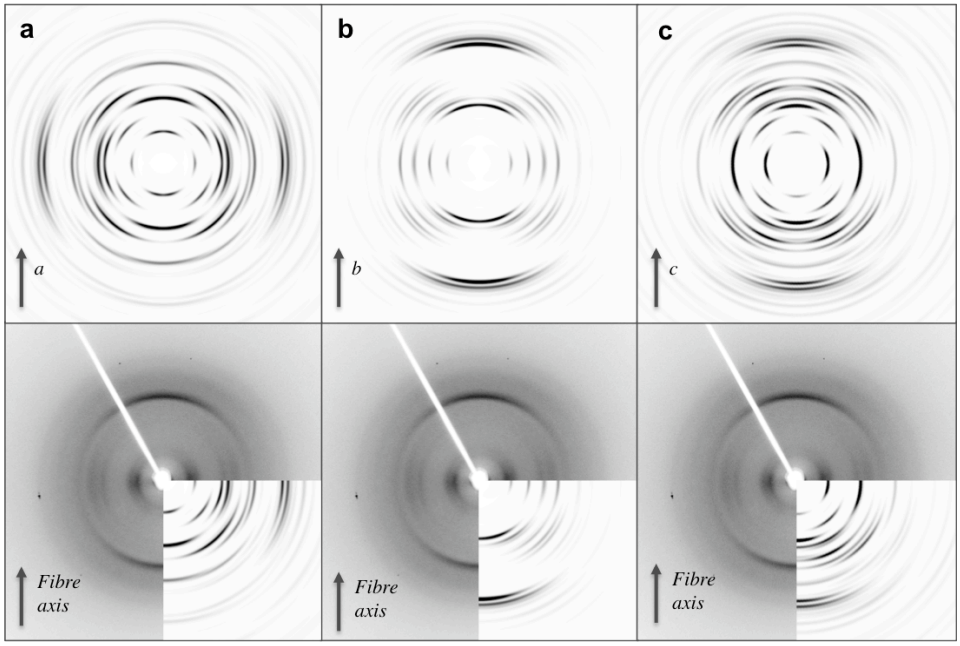


Figure 8.13: Simulated and experimental fibre diffraction of 2-AV assuming each axis of the crystal lattice could represent the fibre axis.

1-GA and 1-AG crystal structure and prediction

Studies into the structure of a related dipeptide not included in the library, 1-GA, gives an opportunity to further compare a crystallisable systems structure to its structure in the gel phase (Adams et al., 2010a). The fibre diffraction data collected from 1-GA in the gel phase is shown in Figure 8.14a. The large number of relatively long-range equatorial reflections arising from the lateral packing of the dipeptide is similar to previously collected fibre diffraction data in that it indicates long-range lateral repeats and may be indexed to unit cells reflecting this as shown in Table 8.9. The large unit cells give rise to the same difficulty in modelling as described in Chapter 7 and previous sections of this chapter. Adams *et al.* report not only on the crystallisation of this, but also, a related 1-position naphthalene conjugated system (1-AG), making attempts to use crystal prediction to predict the structure of the crystal and fibre phase structure (Adams et al., 2010a).

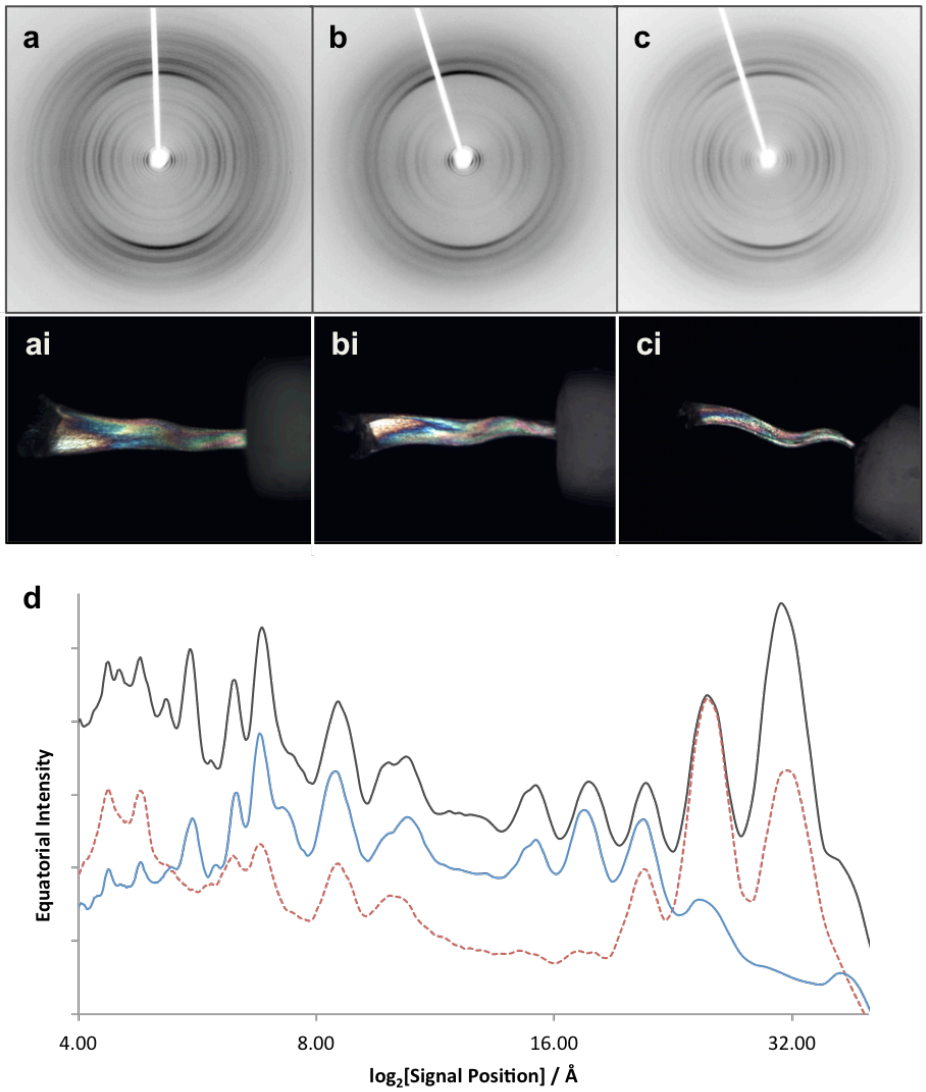


Figure 8.14: X-ray fibre diffraction from gels formed from 1-GA after alignment (a), followed by extensive dehydration by vacuum desiccation (b) and rehydration (c). Corresponding pictures show the birefringent alignments under cross polarisation. The 1D trace shows the changes in equatorial reflections with de/rehydration (d). Grey solid - a, red dotted - b and blue solid - c. The x axis is logarithmic for clarity and intensities have been shifted for comparison (see Appendix III-xiii). All fibre axes are vertical.

Initially it was necessary to establish whether water should be included in structure prediction searches; that is whether water involvement is on a molecular (within the unit cell) or macromolecular scale (between fibrils). The effect of hydration on the structure of fibrillar 1-GA was investigated by XRFD in the normal, a dehydrated and rehydrated state as shown in Figure 8.14a-c. The dehydrated state of the alignment was achieved by a thorough thirteen-day vacuum desiccation and rehydrated by two-minute submersion in water.

System: 1-GA				
	a	*b*	*c*	$\alpha=\beta=\gamma$
Search width:	**40**	**34**	**4.76**	**90**
Predicted unit cell:	**35.19**	**26.51**	**4.76**	**90**
	Miller Indices			
Empirical	**h**	**k**	**l**	**Calculated**
32	1	0	0	35.19
25.9	0	1	0	26.51
21.4	1	1	0	21.17
18.1	2	0	0	17.59
15.4	2	1	0	14.66
10.5	2	2	0	10.59
8.6	1	3	0	8.57
6.8	5	1	0	6.8
6.2	2	4	0	6.2
5.5	5	3	0	5.5

Table 8.9: The lowest error unit cell indexing of the 1-GA XRFD patterns.

This process was found to not grossly affect the visual appearance of the alignment, nor its birefringent properties (see Figure 8.14ai, bi and ci). However water almost certainly participates in the lateral packing structure since extensive dehydration leads to a clear change in the equatorial signals in the fibre diffraction pattern (b). Close inspection of the equatorial signal intensities (d) reveals the signal positions in the dehydrated sample remain the same but their relative intensity is modulated (see Appendix III-xiii), seen largely in the 11 – 19 Å region. The major meridional reflection at 4.76 Å (+0.03/-0.03 Å) is unaltered indicating that the repetitive spacing along the fibre axis is stable despite dehydration. Remarkably, the modulated equatorial reflections can be almost entirely restored upon rehydration (c and d) whilst the major meridional reflections remain constant (see Appendix III-xii) giving strong indication that the structures are not permanently damaged by the dehydration process but in some manner subtly reorganised.

In answer to the initial question of whether water is molecular or macromolecular; these findings are consistent the water being macromolecular. If hydration was integral to the local structure of the unit cell then upon dehydration either a change in the unit cell packing and thus dimensions or complete structural destruction would be expected, in contrast to what is

observed here. The observation of restoration of equatorial reflections upon rehydration further corroborates that the structure is not destroyed and simply modulated in some way. Where the changes occur only on the equator, the structural changes correspond to dimensions perpendicular to the fibre axis. Where they can be restored, this shows a non-permanent structural change. The data is consistent with a model where fibril packing is subtly reorganised upon dehydration due to water playing a role in the macromolecular organisation of fibrils perpendicular to their long axes perhaps in the form of a 'hydration jacket' (Adams et al., 2010a).

Indeed the structure of 1-GA in the crystalline phase is found to assemble in a repeating lattice lacking molecular water (Adams et al., 2010a) as shown in Figure 8.15. The structure is stabilised by an elaborate three-dimensional network of hydrogen bonding, allowing it to propagate in all dimensions distinct from a fibrillar system where hydrogen bonding would be expected to be one-dimensional (Dastidar, 2008). The complexity of the structure can be illustrated by finding that the dipeptide pairs align to four separate planes in three-dimensional space (see Appendix III-xiv) and thus it is clear that the structural elements of assembled 1-GA are not confined to discrete axes as might be expected for a anisotropic structure. The structure cannot be traditionally classified as being in a ß-strand conformation where the first dihedral angles of Φ -169.4 – -168.1 and Ψ -170.7 – 171.1 lie outside of Ramachandran favoured regions (a). This indicates that ultimately the dipeptides cannot be wholly described by the classification used for protein structures and is a reminder that they are small molecules. This shows that, the structure of 1-GA in the crystalline phase is not constrained by peptide backbone conformational interactions for stability (b) but rather would be better described as having complex packing (c) presumed to be the most energetically stable arrangement.

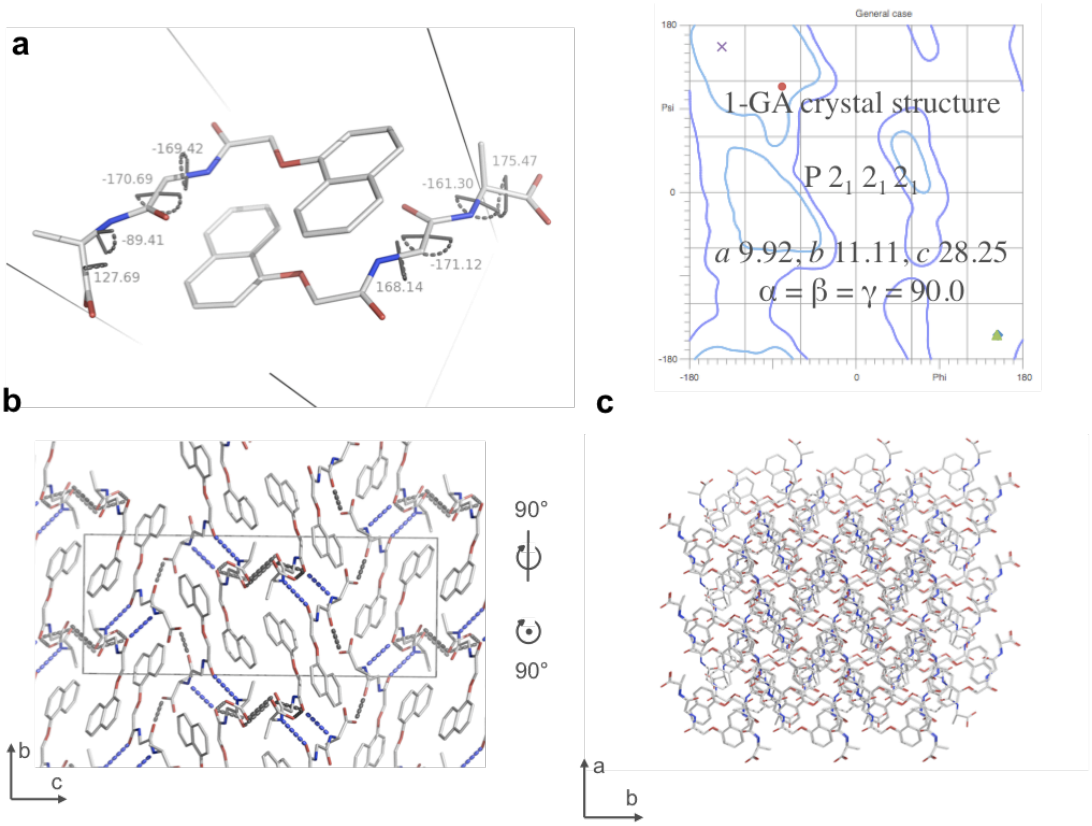
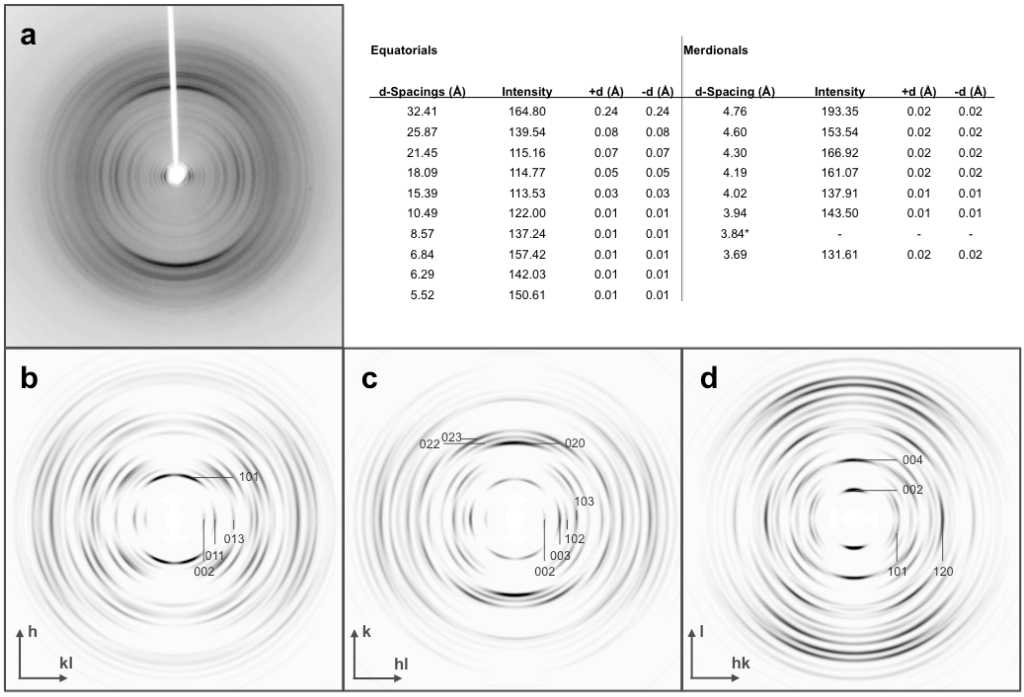


Figure 8.15: The structure of 1-GA in the crystalline phase. The conformation of the dipeptide is nonstandard with the first dihedral angles of the two molecules of the asymmetric unit lying outside of Ramachandran allowed regions (a). The structure is stabilised by a three-dimensional hydrogen-bonding network involving intercarboxyl-carbonyl (grey dashed lines) and interamide (blue dashed lines) interactions (b and c). Coordinates provided by Dr. D. J. Adams (University of Liverpool) and solved by Dr. J. Bacsa (University of Liverpool). Graphics generated in PyMol (DeLano, 2002).

Figure 8.16 shows the simulated fibre diffraction from the crystal structure further illustrating that the crystal structure is not directly related to the structure of the fibre. Evidently, even more so than 2-AA and 2-AV, the complex three-dimensional architecture is reflected in the complexity of the simulated fibre diffraction patterns. By comparison, the experimental pattern of 1-GA exhibits reflections confined to the meridional and equatorial axes implying the repetitive spacings within the fibrillar phase are predominantly aligned to the axes of the unit cell. Simulation from the cell only produces a pattern with reflections, for the majority, confined to meridional and equatorial axes. Including the structure modulates the intensity of the off-axis reflections where the structure contains off axis repetitive spacings (see Appendix III-xv), as seen with 2-AA.



Equatorials

d-Spacings (Å)	Intensity	+d (Å)	-d (Å)
32.41	164.80	0.24	0.24
25.87	139.54	0.08	0.08
21.45	115.16	0.07	0.07
18.09	114.77	0.05	0.05
15.39	113.53	0.03	0.03
10.49	122.00	0.01	0.01
8.57	137.24	0.01	0.01
6.84	157.42	0.01	0.01
6.29	142.03	0.01	0.01
5.52	150.61	0.01	0.01

Merdionals

d-Spacing (Å)	Intensity	+d (Å)	-d (Å)
4.76	193.35	0.02	0.02
4.60	153.54	0.02	0.02
4.30	166.92	0.02	0.02
4.19	161.07	0.02	0.02
4.02	137.91	0.01	0.01
3.94	143.50	0.01	0.01
3.84*	-	-	-
3.69	131.61	0.02	0.02

a 9.92, *b* 11.11, *c* 28.25, $\alpha = \beta = \gamma = 90.0$

Index [hkl]	Position / Å
[002]	14.12
[011]	10.34
[101]	9.36
[013]	7.18

Index [hkl]	Position / Å
[002]	14.12
[003]	9.42
[102]	8.12
[103]	6.83
[020]	5.56
[022]	5.17
[023]	4.79

Index [hkl]	Position / Å
[002]	14.12
[101]	9.36
[004]	7.06
[120]	4.85

Figure 8.16: Simulated fibre diffraction from the crystal structure of 1-GA. Fibre axis a / 100 (b), b / 010 (c) and c / 001 (d). All fibre axes are vertical.

A modification to the sequence of this dipeptide, to 1-AG, produces a system with a drastically altered assembly architecture and phase preference for crystallisation (see Appendix III-xvi). The peptide backbone, like 1-GA, is observed to have one of its dihedral angles outside of Ramachandran allowed regions, indicating the need for efficient packing over peptide conformational stabilising interactions for lattice stability. However, the structure of 1-AG in the crystalline phase is stabilised by intra and intercarboxyl-carbonyl and also interamide backbone interactions forming a two-dimensional hydrogen-bonding network unlike 1-GA crystals that have a three-dimensional network. It might be expected that a two-dimensional bonding network has a higher propensity for anisotropic growth/assembly over a three-dimensional bonding network but despite this the system preferentially forms crystals. Where

the 1-AG system does not form a gel phase on which XRFD can be used it is not appropriate to perform simulated fibre diffraction; although why it does not form fibres is likely significant.

Perhaps the 1-GA crystal structure occupies an energetic minima and the fibrillar phase is in a trapped less energetically favourable state. The same has been suggested for amyloid self-assemblies (Hwang et al., 2004). This point was addressed where Adams *et al.* explored the crystal energy landscape of 1-GA and 1-AG (Adams et al., 2010a). It was found that the crystal structure of 1-AG could be successfully predicted but the search was limited to unit cells containing one asymmetric unit per unit cell and so a prediction for 1-GA was not made where the crystal structure unit cell contains two asymmetric units (Adams et al., 2010a). Despite this, the crystal predictions for 1-GA revealed two packing modes that may help understand the gel structure of 1-GA. The packing modes are described as compact-columns and open-tapes and are shown in Figure 8.17 (Adams et al., 2010a). Both are stabilised by a two-dimensional hydrogen-bonding network with interamide backbone interactions, though perpendicular to the amide hydrogen bonding the compact-column differs by possessing intercarboxyl-carbonyl interactions and the open-tape intercarboxyl-carboxyl interactions. The compact-column has two backbone amide interactions between the second amide carbonyl oxygen and the two amide nitrogen atoms, whereas the open-tape has two unique backbone amide interactions between unique amide nitrogen and carbonyl oxygen atoms. Of the architectures shown in Figure 8.17 the open-tape structure is less stable than the compact-column (Adams et al., 2010a) structure perhaps indicating that efficient packing is energetically favoured over increased hydrogen bonding.

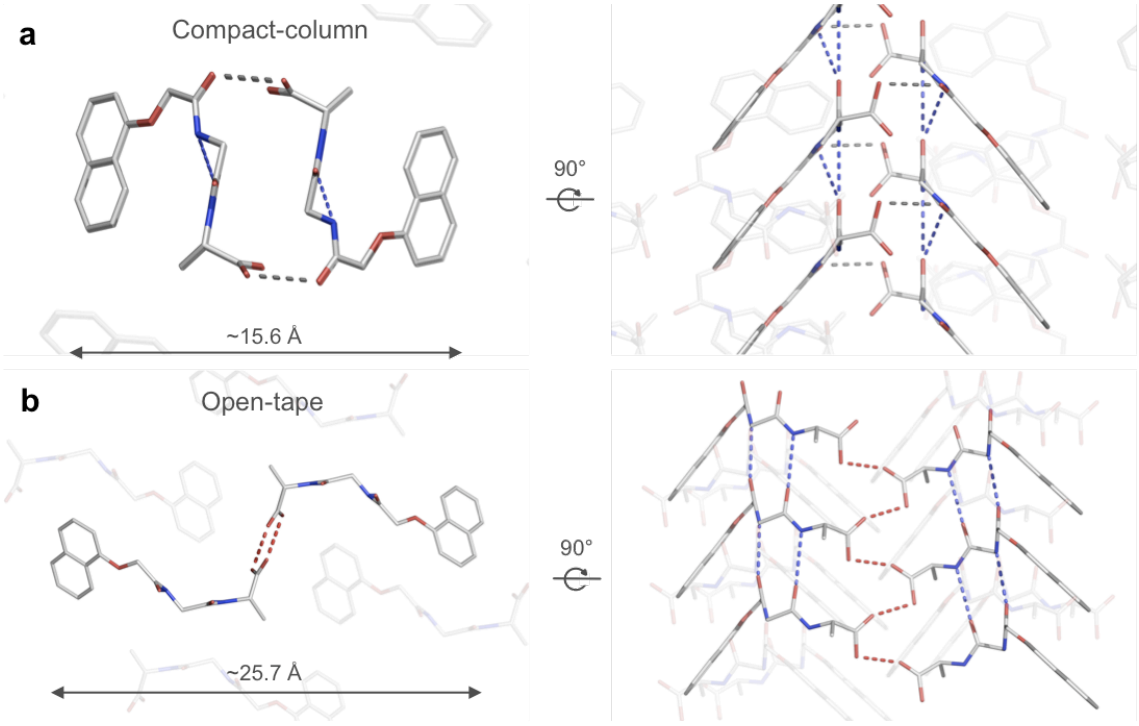


Figure 8.17: The two architectures identified in attempts to predict the structure of 1-GA. The compact-column (a) architecture represents a -172.28 kJmol^{-1} structure stabilised by intercarboxyl-carbonyl hydrogen bonding (grey dotted lines) and the open-tape (b) architecture represents a -171.83 kJmol^{-1} structure stabilised by intercarboxyl-carboxyl hydrogen bonding (red dotted lines). Both exhibit interamide hydrogen bonding (blue dotted lines). No interactions were found between dipeptide pairs. Adapted from (Adams et al., 2010a). Graphics generated in PyMol (DeLano, 2002).

The authors note that the open-tape packing mode may be less likely to support one-dimensional growth due to a steric requirement for laterally associated open-tapes (Adams et al., 2010a) but in contrast it is tempting to believe that the two parallel interamide hydrogen bonds (see Figure 8.17b) would result in a greater propensity for anisotropic one-dimensional growth than the single interamide hydrogen bonding of the compact columns. On the latter basis, the open-tape packing mode may well better represent the fibrillar structure and the compact-column the crystal structure. Indeed, although the intercarboxyl-carbonyl hydrogen bonds are only formed on one side of the dipeptide pairs, the crystal structure of 1-AG is akin to the compact-column architecture.

Although the crystal structure of 1-GA does not appear to represent the fibre phase, perhaps the predicted compact-column or open-tape architecture do. Predicted structures of 1-GA in both the compact-column and open-tape architecture were compared to the experimental fibre diffraction data but where the unit cells of the predictions by Adams *et al.* (see Appendix III-xvii) are not obviously related to the predicted unit cell from the fibre diffraction pattern shown in Table 8.9 it is very difficult to experimentally validate which packing mode best represents the fibrillar structure (Adams et al., 2010a). Fibre diffraction simulations were performed on the two packing modes shown in Figure 8.17 (see Appendix III-xviii) but no correlation could be found to the experimental fibre diffraction pattern. This is not necessarily because neither packing mode is correct but because the reflections positions in the simulation are governed by the unit cell dimensions and are simply not possibly indexed to the fibre diffraction pattern reflections. Where the unit cell in the fibre phase is likely large, the relatively small cells in the 1-GA predictions may represent a sub-lattice within the 1-GA fibrillar structures. It would therefore be beneficial to computationally explore the compact-column and open-tape packing modes within the predicted unit cell from the 1-GA fibre diffraction.

Crystallisation represents a phase separation event that leads to three-dimensional structure formation and thus no fibre growth and gelation, the determinants of which are notably complex. Why crystals are preferentially formed over gel fibres remains an open question.

8.2.10 Pre-gelation assembly, morphology and structure

Further corroborating the discrete phenomena of self-assembly and phase preference are the observations of anisotropic material in diphenylalanine dipeptide solutions prior to gelation as discussed in Section 8.2.1 (see Appendix III-i). This was investigated by the characterisation of this material as shown in Figure 8.18. Assuming that the once dried to the grid surface the acidic staining conditions do not affect the dipeptide, the diphenylalanine containing dipeptides have a fibrillar appearance (although this has been attributed to worm-like micelles (Chen et al., 2011)) prior to gelation explaining the increased viscosity upon pipetting. Their observation at high pH suggests that the diphenylalanine motif is able to promote fibrillation by overcoming the favourable solvation energies associated with the carboxylate terminus. However, where these systems are not true gels at high pH, the mechanism of their assembly must be distinct from the mechanism of gelation. This might be the case but assumes that the self-assembly mechanism at high pH is the same as that at low pH. It has recently been shown that the high pH fibres of 2-FF have a structure consistent with worm-like micelles (Chen et al., 2011). To investigate the structure of the fibrillar material at high pH fibre alignments were prepared, but only CN-FF was found to align. The fibre diffraction from this alignment is shown in Figure 8.18. The pattern is indicative of a high degree of order in the pre-gelation assemblies and compared with the low pH fibre diffraction from CN-FF (Figure 8.8), exhibits less radial averaging where the fibres are better aligned, perhaps related to a greater capacity to align fibrils when not in the gel cross-linked state. Of particular interest though is that the diffraction from the high- and low-pH fibres is equivalent, showing that the structure of the fibres prior to gelation is already established. In this unique case the self-assembly mechanism, probably driven by the diphenylalanine motif, leading to anisotropic fibre growth is distinct from the secondary gelation mechanism induced by a pH decrease.

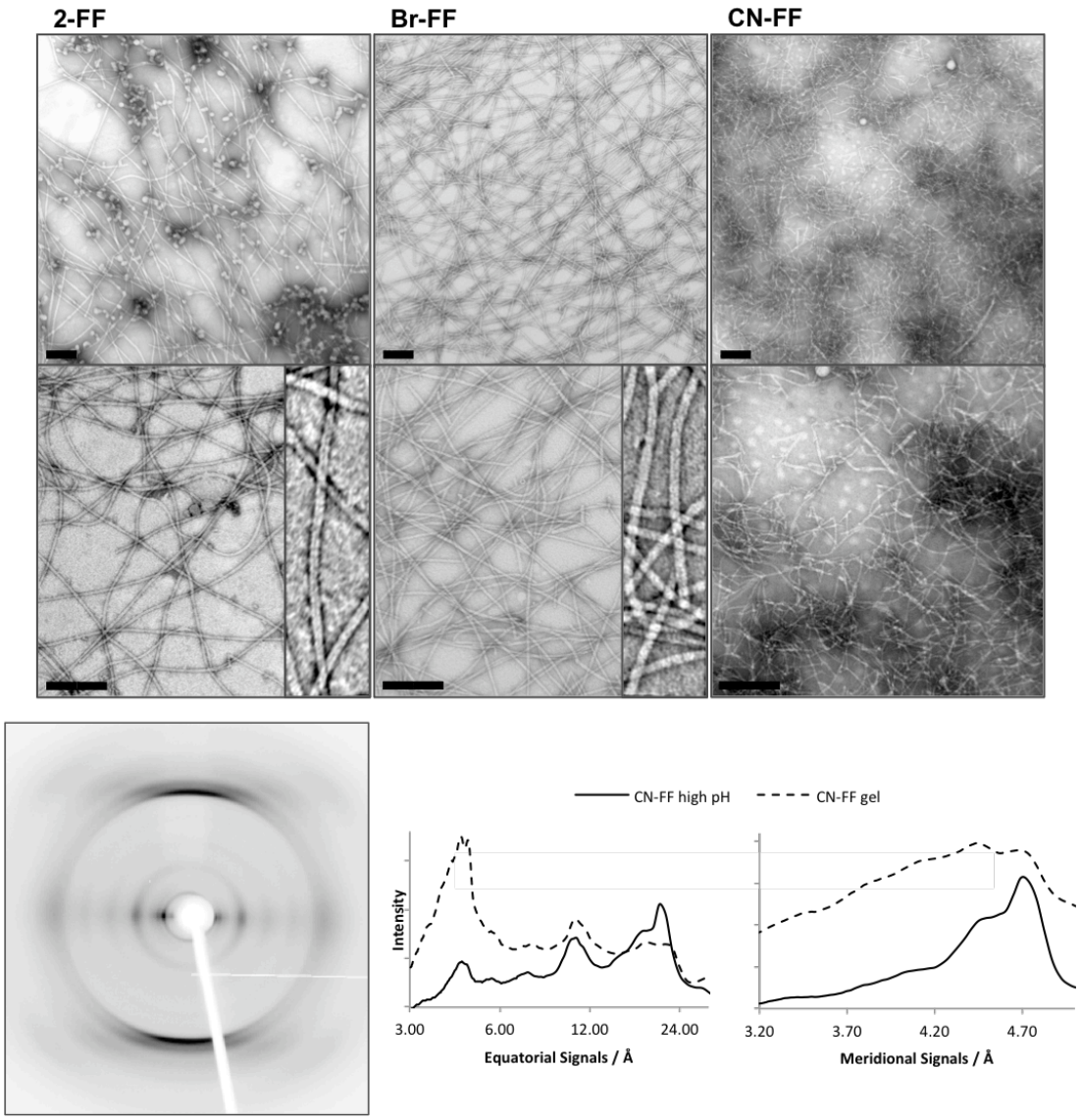


Figure 8.18: Transmission electron micrographs of diphenylalanine dipeptides prepared at high pH and X-ray fibre diffraction exhibited by an alignment of CN-FF. The scale bars represent 200 nm. The inserts show FFT crops illustrating morphological features – not to scale. The fibre axis in the XRFD pattern in vertical.

8.2.11 Structural insights from gel mixtures

This work has tried to link structural features to material properties, based on the thorough characterisations already established this was further explored through dipeptide system mixing. It was hypothesised that if two systems with similar or differing pKa values (see Table 8.1a-b) were mixed, upon gelation the resulting fibres might co-assemble and self-sort respectively (see Figure 8.19a & b) and if the resulting materials properties differed then a structural basis might be identified. The former co-assembled scenario has not yet been observed but by mixing 2-AA and Br-AV in a 1:1 ratio, upon assembly a self-sorted mixture of fibrils is observed as indicated by the XRFD exhibited from the mixed gel. As shown in Figure 8.19c-e the mixed fibres (c) are a simple overlay of the independent patterns of 2-AA (d) and Br-AV (e) indicating that no new structures are formed upon mixed self-assembly and gelation. The equatorial and meridional traces may be inspected graphically supporting the evidence for overlaid diffraction patterns. Further, the signals expected to arise from this phenomena may be calculated from the individual diffraction pattern of the dipeptide independently and found to correlate very well with the experimental mixed pattern down to a resolution of ~20 Å, perhaps implying differences in the long-range lateral packing of the self-sorted fibres mixtures.

Self-sorted fibres have the same local structure (<20 Å) as their independent assemblies but subtle differences >20 Å up to the length scales that this fibre diffraction is able to show (~40 Å). The newly formed gel might have been expected to have differing rheological properties but no difference is detectable with the rheological methods usually employed to characterise these materials (Dr. D. J. Adams, verbal communication). Perhaps co-assembled mixtures of dipeptides would result in differing material properties, however this is found not to be the case, although the observations here are no less interesting. Efforts will continue to try and achieve co-assembly where a structural difference may be found to relate to material property changes.

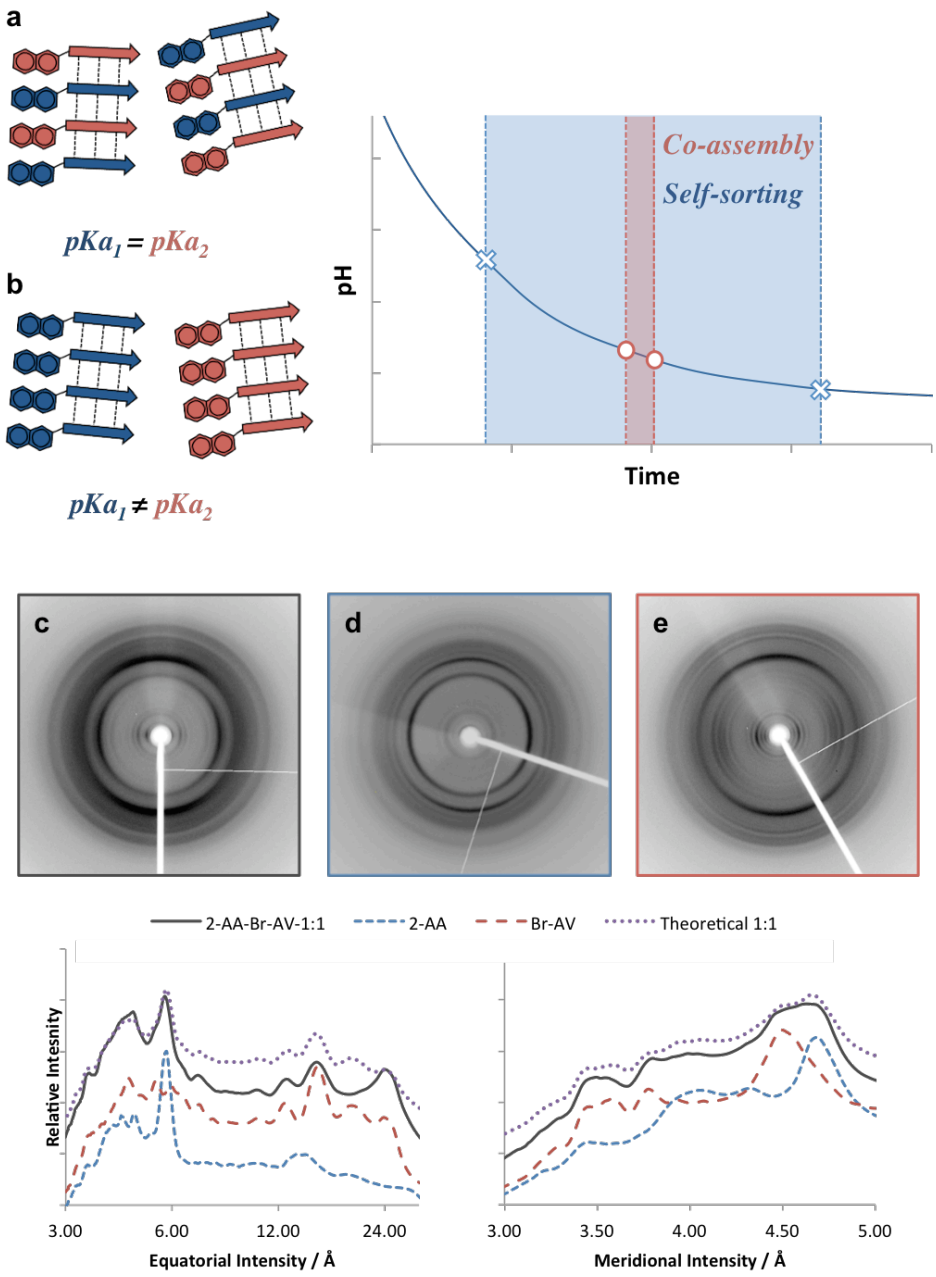


Figure 8.19: The theoretical basis for self-sorting based on assembly by pH decrease (a-b). Self-sorting is exhibited by the two systems 2-AA (pKa 5.1) and Br-AV (pKa 5.8) when assembled with 8.9 mgmL^{-1} GdL in a 1:1 mix at a total dipeptide concentration of 5 mgmL^{-1}. Mixed fibres are observed to be a simple overlay (c) of the dipeptide fibres individually (d-e) consistent with self-sorting. All fibre axes are vertical.

8.3 Summary

Principally this work has aimed to determine the structure of a library of LMWG in the fibrillar phase. In doing so an attempt has been made to identify what relationship exists between the structures of self-assembled LMWG in the fibrillar phase to the properties of the gels they form. The assembly process has also been examined. By characterising a library of LMWG it was possible to compare the similarities and differences between the molecular and fibrillar structure to give insight into the effect of LMWG modification on fibrillar structure and gel properties. Within the library there were systems known to enter different phases (i.e. fibrillar gels, turbid gels and crystals) (Chen et al., 2010c) and so this characterisation presented an opportunity to dissect the determinants for these phase preferences. Further, in unique cases LMWG are observed to crystallise from the gel phase. These might be compared to the fibre phase to assist in the determination of the self-assembled fibre phase structure but the results of this analysis raises more fundamental questions regarding whether the crystal structure of these classes of molecules may be used to represent the structure in the fibre phase as is the case for other LMWG reported in the literature (Dastidar, 2008).

The determination of the structures of self-assembled LMWG fibres has been conducted with a combination of techniques. Ultimately, full models representing the fibre phase have not been constructed (as also discussed in Chapter 7) but a number of structural characteristics relating to the assembly and final structure of the LMWG have been identified. There are several levels with which to characterise the material formed by the dipeptide library. The first is the phase that the material assembles into whether that be crystalline, gel or turbid gel. Beyond this, the gels are known to have different mechanical strengths and then as characterised here, morphological and structure differences.

#	C_1	R_1	R_2	Phase†	G' / Pa†	Minimum fibril width / nm	Fibril morphology	Structure	Relative h-bond strength / %	Nap helicity	Nap long axis / °	Nap short axis / °	Orthorhombic unit cell / a b c
1	H	L-Ala	L-Ala	Gel/Crystal	83000	9.3	Thin-filament	ß (A)	79	-	90-54	-	39.91, 14.72, 4.68
2	Br	L-Ala	L-Ala	Gel	28	21.7	Crystalline nanotube	ß	40	R	54	54	65.05, 33.00, 4.30
3	CN	L-Ala	L-Ala	Turbid Gel	79000	4.3	Ultra thin-filament / Helical	ß (A)	0	-	90-54	54	39.89, 20.25, 4.57
4	H	L-Ala	L-Val	Gel/Crystal	n.d.	17.1	Thick-filament	ß	50	-	90-54	54	35.39, 15.24, 4.70
5	Br	L-Ala	L-Val	Gel	25000	17.2	Thick-filament / Helical	ß (A)	79	L	90-54	54-0	65.66, 48.20, 4.51
6	CN	L-Ala	L-Val	Gel/Crystal	n.d.	14.8	Helical	ß	4	-	90-54	90-54	46.93, 14.78, 4.54
7	H	L-Phe	L-Phe	Gel	72000	12.7	Thin- / Thick-filament	ß (A)	65	L	-	-	44.66, 34.44, 4.67
8	Br	L-Phe	L-Phe	Gel	180000	9.3	Thin-filament	ß (A)	100	L	-	-	66.00, 28.34, 4.71
9	CN	L-Phe	L-Phe	Gel	15000	9.5	Thin- / thick-filament	ß (A)	79	-	-	-	46.84, 12.49, 4.54
10	H	L-Phe	L-Val	Turbid Gel	2000000	8.5	Thin-filament	ß	93	L	90-54	54	x, 11.30, 4.70
11	Br	L-Phe	L-Val	Turbid Gel	100	11.3	Thin-filament / Nanotube	ß (A)	79	R	90-54	54	68.62, 42.64, 4.66
12	CN	L-Phe	L-Val	Turbid Gel	29000	22.6	Helical-filament / Nanotube	ß (A)	93	R	54-0	54	46.49, 14.32, 4.64

*Table 8.10: A table summarising the phase preference and material properties of the dipeptide library compared to the structural information determined. *FTIR data collected and provided by Dr. L. Chen (University of Liverpool), possible antiparallel-ß interpretations are made by parenthesis.*

The material properties of the gels may be concisely defined by their G' values as summarised again in Table 8.10 along with the structural features characterised. Simple material observations, such as the formation of clear or turbid gels, can be described that might be relatable to molecular and structural features. However, no clear distinction can be made between the turbid or clear gel forming systems. Where all of the systems are found to adopt a self-assembled fibrillar structure, in the gel phase, it is particularly interesting that diphenylalanine dipeptides are self-assembled prior to gelation. Characterisation of the structure of one of the pre-gelation self-assembled structures finds it to be the same as the assemblies in the gel phase, suggesting that self-assembly is in some cases distinct from gelation. This is consistent with observations made on Fmoc-Phe-Phe. It has been noted that the ~1720 cm^{-1} FTIR peak is attributable to the –COOH group, as would be expected for dipeptide systems that assemble after protonation of this group (Chen et al., 2010a). Where this peak is absent from the FTIR spectrum of gels of Fmoc-Phe-Phe this indicates that the carboxyl group is not protonated (Smith et al., 2008, Tang et al., 2009) and thus this system is likely not assembling via the protonation mechanism but like the diphenylalanine dipeptides reported here is able to assemble

due to the stabilising interactions of this motif. Otherwise each system is found to self-assemble into fibrillar structures concurrently with gelation. The substitutions made to the naphthyl group dramatically alter its spectral properties, which has been characterised and utilised in CD measurements to, where possible, follow the self-assembly of these systems. Each exhibits complex spectra that cannot be interpreted like normal CD spectra but indicate a structuring of the dipeptide backbone consistent with ß-sheet formation and ordering of the naphthyl groups into a chiral environment as is also reported in Chapter 7. The adoption of a ß-sheet conformation is comparable as characterised by CD, but the exact architecture i.e. parallel or antiparallel is variable between systems when characterised by FTIR. There is evidence of exciton coupling implying π-π aromatic stacking, which is apparently unaffected by the substitution of the naphthyl 6-position. The chiral and geometrical arrangement of the naphthyl groups have been determined in systems where possible. Although the peptide configuration across the library is comparable, the chiral arrangement of naphthyl groups in some cases appears related to the dipeptide sequence but this is not consistent across the library. Further details about the geometrical arrangement of the naphthyl groups indicates the majority align by their long-axes in a face-to-face arrangement between 90.0 – 54.7° as found previously for Br-AV (see Chapter 7), although the short-axes naphthalene alignments are more complex. By these observations, although the same dipeptide conformation and π-π stacking is adopted by each system, evidently the precise architectures of these molecules in the fibrillar phase is complex and not simply related to the molecular structure.

The unit cells into which these complex architectural arrangements pack have been determined by XRFD. There appears to be a correlation between the naphthalene conjugation and the separation of ß-strands along the fibre axis as well as the local and long-range perpendicular packing within the fibrils. The precise ß-strand separation appears to be modulated by the nature of the naphthalene conjugate whereby bromo and cyanonaphthalene conjugates lead to a shortening of this distance. However, the major repetitive spacing along and perpendicular to the fibre axis are found to vary less among the aromatic dipeptides possibly owing to the

phenylalanine residues causing preferential dipeptide assembly. More dramatically affected are the lateral packing interactions of these ß-sheets, the long-range order of the determined unit cells appears to be entirely dependent on the naphthalene conjugation. By this analysis, the unit cells are found to be large relative to the dipeptide molecule, containing between 6 – 39 asymmetric units, but are found to be related by the nature of the naphthalene conjugate C_6 (H, Br or CN) or aromatic residue content. Where these cells have such long-range order it is tempting to consider that long-range interactions might govern fibril lateral assembly and in turn affect gel properties. Indeed, investigations on the hydration of 1-GA find variability in the long-range packing modes of these molecules, pointing to some role for this type of interaction. It might thus be expected that where the dipeptide unit cells may be grouped by naphthalene conjugate, the material properties could also be grouped by naphthalene conjugate, however this is not the case.

Clearly the link between molecular structure, assembly architecture and the relationship of these to the lattice that makes of the self-assembled fibre is complex. To further attempt to elucidate this relationship an analysis of crystal structures of the LMWG was made. Hydrogen bonding networks were identified in the crystal structures of 2-AA and 2-AV that are predominantly one-dimensional perhaps consistent with fibre formation. However perpendicular interactions, can be found, that are likely promoting crystal growth and an absolute comparison to the fibre diffraction data finds the structures are not representative of the fibre phase. The idea that they are perhaps subtly related was pursued and finds that although the lattice packing in the fibre may be less condensed than the crystal, the internal structural spacings and thus stabilising interactions of the crystal may be present in the fibre, but if so, are reorganised. These observations are consistent with the differences in crystal versus fibre formation, whereby the interactions and interatomic separations within a fibre are known to be aligned, thus promoting one-dimensional growth whereas in the crystal structure they are aligned to multiple axes allowing three-dimensional growth. Differences between crystal and fibre structure has been noted in other investigations into systems that can adopt either state (Marshall and Serpell,

2010) as appears to be the case here. However, considering the similarities, it is tempting to think of the crystal structure at least to be structurally related to the fibre phase structure although the absolute geometrical configuration is different.

Where comparisons are made in some systems, they cannot be made in others; the experimentally determined crystal structure of 1-GA was also found to not represent the fibre phase. This is consistent with attempts made by Adams *et al.* to predict the structures of the molecules 1-GA and 1-AG in the fibre and crystal phase (Adams et al., 2010a). Indeed, it was found the method developed was able to successfully predict the structure in the crystal phase but not the fibre phase. The predictive methods employed by Adams *et al.* were able to identify various packing modes that may grant further insight into the different architectures underlying crystal or fibre structures but further computational work is required to consider these in the context of the unit cells indexed for the dipeptide systems here.

The determinants of crystallisation must be carefully considered in the same manner as the determinants of self-assembly where the latter also appears to be a distinct mechanism from gel formation. Equally so, the types of gelation, that is to form a clear or turbid gel, as well as the mechanical strength of the gel formed is not based on a straightforward relationship between molecule identity or self-assembled structure. The basis for hydrogelation and corresponding strength is well known to be based on fibril cross-linking (Lin, 2003) and that may be the case but the link between self-assembly, fibril structure and fibril cross-linking is not immediately clear here. Further, based on this understanding, the self-assembly of a LMWG gelator must be a prerequisite for gelation to occur. However, the observation of material that is fibrillar in appearance and in one case has a comparable self-assembled structure pre-gelation, it must be concluded that the self-assembly of peptidic LMWG does not necessarily lead to gelation. Clearly there is a distinction to be made between self-assembly propensity and gel formation. The basis for self-assembly seems to be clear in that molecules should be designed to have favourable one-dimensional interactions however the prediction of how these assemblies will

then interact to reproducibly form materials of specified properties is not clear. It might be expected that the differences in macromolecular properties (i.e. to crystallise, gel or form a turbid gel) would be underpinned by a difference in molecular or self-assembled structure, but there are no obvious trends in the biophysical or structural characterisation that make these systems distinct.

Where other techniques find similarities between two dipeptides apparently unrelated with regard to their phase preference or gel mechanical strength, it is unclear how local packing is linked to gel properties. Studies on the gel phase, mixtures of systems, as well as hydration studies have inferred significance in the long-range lateral packing of fibrils. Perhaps long-range lateral packing of fibrils governs the macromolecular properties of hydrogels. Future work will need to consider not the self-assembly mechanism and structure but higher order assemblies of these. The collection of cryo-electron microscopy (cryo-EM) data may permit docking to the crystal structures to better dissect the relationship between the short- and long-range interactions of these systems but it should be kept in mind that the crystal structures do not fully represent the fibre phase. Solid state nuclear magnetic resonance spectroscopy may prove useful to probe the fibre phase and produce accurate models for cryo-EM docking but this would also provide an experimental framework to apply the constraints provided by the exciton coupling theories discussed and consider their context within the repeating lattice of the fibre. Information from linear dichroism, with careful consideration of the alignment method, would also be very beneficial to add to the biophysical data presented here. By these additional works and building on the foundation of information here it may be possible to fully determine the link between molecular structure, fibrillar architecture, fibre formation, phase preference and material properties. Subsequently, LMWG design will be able to consider self-assembly and gelation individually and tailor the propensity for these two separate phenomena to the desired application.

9 Concluding Discussion

The work presented here has described the self-assembly process and final structure of a range of small self-assembling peptidic systems from naturally derived to designer. In their characterisation, the length scales of self-assembly have been described from tens of Ångstroms to hundreds, illustrating the elegant route to large structure formation these systems present. Whether relevant to disease or technology, the methodologies used to best determine the self-assembly and structure in the fibrillar phase has again illustrated the broader implications of studying anisotropic materials. A great deal of information has been uncovered on the systems presented here, but also highlighted are some outstanding and enticing questions outlined as described.

9.1 Waltz

Investigations on the structure of the Waltz peptides has identified three structures that represent the structure of these systems in the fibrillar phase. Fibre diffraction is well established as being able to identify the cross-ß architecture and indeed has identified the three Waltz peptides investigated here to adopt this structure, with a high degree of order in one case. Further, in principle X-ray fibre diffraction is able to distinguish between precise architectural arrangement and even side chain conformation, though this relies on a system under investigation to have a high degree of crystallinity. Thus, the structures identified here best represent the architecture and side chain conformation of these peptides in the fibrillar phase. Though, it should be noted that the difference between the simulated patterns from the modelled structures are slight. This may indicate that the fibres contain a number of morphologies, as observed for other amyloidogenic systems (Nelson et al., 2005, Sawaya et al., 2007), where several different models are representative of the fibrillar phase. Undoubtedly, the arrangements differ between the Waltz peptides investigated, which is surprising given position sequence similarity (Position

3 – Tyr or Phe , position 5 – Ile) of the systems, however this only further reflects this conformations adaptability.

The identification of the ordered cross-ß architecture adopted by these systems confirms the Waltz algorithm's ability to identify ordered aggregation as reported (Maurer-Stroh et al., 2010). In addition, as the development of future versions of Waltz requires more structural information the structures here can be incorporated in this process.

9.2 αSß1

The elaborate supramolecular organisation of a fragment of α-synuclein has been shown. Where the long-range assembly of small peptidic systems requires further investigation in other systems it has directly been observed here. The elaborate arrangement of the cross-ß architecture on previously unobserved length-scales produces a nanotubular morphology. Though the two levels of structure, molecular and supramolecular, can be modelled individually, this work has only touched on the connection between the two. This calls for future work to address this connection; where here it is beautifully observed, in other investigations it might prove essential to understanding the systems behaviour.

9.3 Dipeptide low-molecular weight gelators

A large range of peptidic low-molecular weight gelators (LMWGs) have been reported in the literature (Adams et al., 2010b, Chen et al., 2010a, Chen et al., 2010c, Smith et al., 2008, Xu et al., 2007). The principles and mechanisms of their self-assembly are as varied as the precise molecular structure of the LMWG themselves. The characterisation of the self-assembly and structure of the naphthalene conjugated dipeptide Br-AV is the most comprehensive to date. It illustrates the high degree of reproducible control available over the self-assembly of a LMWG through the protonation mechanism. In line with this, the biophysical and structural

investigations on the course of this process have been analysed and compared to reveal how precisely the self-assembly of this molecules proceeds. At high pH, Br-AV is reported to have a micelle structure above the critical micelle structure (Chen et al., 2011), although, as measured by circular dichroism the dipeptide molecules are unordered and naphthyl groups not in close association. Evidently, as the pH of the solution in which the dipeptide is dissolved decreases, the molecules in solution enter into close proximity such that π-π interactions giving rise to exciton coupling and ordering of the dipeptide backbone may occur. The measurements made here and elsewhere (Chen et al., 2010a) indicate the development of ß-sheets. As far as the investigations here can probe, these events happen concurrently. This process leads to one-dimensional self-assembly as evident by the observation of fibrillar material.

The structures of these self-assembled fibrils may be probed by XRFD to determine the repetitive interatomic separations parallel and perpendicular to the fibre axis. The data are consistent with the formation of an architecture that may be likened to the cross-ß architecture of amyloid self-assemblies. In this case, the ß-strands are spaced at 4.51 Å and the ordered lateral association of these structures is observed up to ~70 Å. The long-range order exhibited in the fibre diffraction is also reconcilable with the complex lateral association observed in the fibrils themselves by transmission electron microscopy (TEM). Based on the XRFD and TEM observations, over the course of self-assembly, a model may be proposed whereby the structure of the fibres is adopted rapidly. As gelation proceeds fibril entanglement increases and becomes harder to overcome, perhaps contributing to the gels rigid behaviour of the gels at low pH.

The structural architectures may be described, but are ultimately very difficult to validate where the full scale over which order occurs is larger than the range over which it is feasible to reliably model with the information currently available. Nevertheless, the cross-ß architecture is adopted with an antiparallel stacking of ß-strands at 4.51 Å. Perpendicular to the fibre axis the data are consistent with ß-sheets being separated by 6.02 Å. This brings the naphthyl groups in close proximity to stack with their long axes in an approximate face-to-face arrangement which are

tilted not more than 54.7° away from the fibre long axis (Chen et al., 2010a). The hydrogen bonding pattern in the proposed model only involves inter-amide hydrogen bonding, no doubt integral to the directional growth of these assemblies. Although clearly the aromatic interactions in this system are important, it is curious they are not disrupted by the naphthalene conjugation of the 6-position bromine atom.

It would appear that the pH driven protonation mechanism induces self-assembly but distinctly contributes to macromolecular changes that affect gelation. These changes occur over time scales at which fibrillar structure has been probed but no clear structural changes are found to occur with changing macromolecular properties, where the fibre structure is adopted rapidly and remains stable over gelation. It was hypothesised that across a library of dipeptides perhaps the differing material properties of the gels formed could be correlated with structural differences. The same difficulty in structure determination is encountered as with Br-AV but many structural features have been gleaned. All systems are found to be structurally analogous to Br-AV with regard to anisotropy, aromatic interaction and ß-sheet formation and stabilisation. The dipeptides all exhibit XRFD data consistent with the cross-ß architecture though the exact arrangement appears to be related to the nature of the naphthalene conjugate. That said some structural relationships may also be based on the dipeptide sequence.

The relationship between molecular structure and fibrillar self-assembled structure is complex. This in turn makes it difficult to pinpoint a relationship between molecular structure and gel properties. This is confounded by the investigations of systems that enter different material phases, i.e. the formation of turbid gels or crystalline phases. If inspected in isolation the data does not indicate a distinction in the turbid gelators, nor an indication of why one system has a propensity to crystallise over another. Further, the crystalline structures of these systems are brought into question, as later discussed. What is clear, is that the characterisations here are able to describe the self-assembly process better than they can gelation. Indeed, in the studies on

diphenylalanine dipeptides, it would appear as though self-assembly is distinct from gelation, although a necessary precursory stage.

The indications of long-range changes upon dehydration, along with the difficulty in determining order over these scales and the changing long-range lateral association of fibres indicates an importance in these kinds of supramolecular interactions.

9.4 Methodological implications

The results presented in this work are a clear illustration and reminder that there are a number of methodological considerations to be made when working with fibrillar systems. Although the alignment of fibrils is often necessary to make biophysical and structural measurements, the uncontrolled alignment of fibrils can lead to artefactual data in particular in biophysical measurements. LD artefacts in CD measurements are reported in Chapter 5 and 7, there are reports in the literature of checking for the presence of LD artefacts (Andersen et al., 2010) but such a clear example of the presence of such phenomena in CD measurements as observed here has not, at the time of writing, been reported. The cuvette rotation and sonication methodology reported here should routinely been employed in the inspection of fibrillar material by CD (see Chapter 5).

Where alignment can be controlled by Couette flow, LD has been measured in a number of amyloid self-assembled systems (Adachi et al., 2007, Andersen et al., 2010, Dafforn et al., 2004, Marshall et al., 2010) to confirm the cross-ß architecture. The same successful LD measurements are reported in Chapter 5 but the attempt to calculate chromophore orientation is novel at the time of writing. Further work is needed to realise the potential of these measurements but the principles of these calculations are outlined in the attempt made here.

The biophysical and structural techniques presented here are rapid, and accurately represent the fibrillar phase, however the modelling process is fairly laborious. This would benefit from an automated computational workflow. If each of the structure determination stages presented in Chapter 5 could be automated and connected within a single program then fibre diffraction used for the elucidation of models representing amyloid structures could be more readily utilised. Of course, this is still dependent on the collection of high quality X-ray fibre diffraction data where an element of serendipity plays a role in the quality of an alignment of fibres. The crystal structures of amyloidogenic short peptides (Sawaya et al., 2007) have made large contributions to the understanding of this three-dimensional architecture but, if representative, the structures can only describe a snap shot of the structure in the fibrillar phase, as evident in the polymorphism found in the peptide GNNQQNY (Nelson et al., 2005, Sawaya et al., 2007). It is true from the observations made in Chapter 5, and as widely appreciated in the amyloid community, that amyloid fibril structures are polymorphic. It is one of the strengths of X-ray fibre diffraction that this phenomenon can be directly observed, but also one of its weaknesses where data quality will be subject to sample polymorphism. Crystalline scattering techniques avoid this but only report on short-range order whereas X-ray fibre diffraction, can report on the long-range order of a fibrillar system. Again, it thus cannot be stressed enough that the structural descriptions of amyloid must be made from a wide variety of techniques to develop the most comprehensive view of these systems.

The evidence shown here supports a view that long-range order is present in self-assembled systems, at length-scales outside the range of high-resolution techniques, as shown in Chapter 6 – 8. It is enticing purely as a scientific problem to try and describe the connection between short and long-range self-assembly, but I contend that there is an important role for long-range interactions in the determination of the macroscopic properties of fibrillar matrices that constitute hydrogels. The short-range and long-range order of these systems, as observed by the real-time alignment data presented in Chapter 5, does not change over the dehydration process

and so supports that X-ray fibre diffraction is representative of the fibrillar structure in solution and the interatomic repetitive separations and interactions observed here are not artefactual.

The accurate representation of the structure of self-assembled systems has been questioned in reports already in the literature (Marshall et al., 2010) and is again raised in this work. Shown in Chapter 8, three peptidic LMWG in the crystalline phase are shown to not represent the fibrillar phase. Although care should be taken in comparing peptidic LMWG to short peptide systems, the principle that the crystalline phase may not represent the self-assembled fibrillar phase should be kept in mind. That is not to say that the two phases may not be subtly related. Clearly, in a large polypeptide the stabilising three-dimensional interactions are many and not as easily influenced by the forces that govern crystallisation. However, where crystallisation is essentially a self-assembly process, this could have a dramatic influence on the forces governing a self-assembled fibre when induced to crystallise. Simulated fibre diffraction can be used to investigate this and can rapidly identify if there are fundamental differences between these two phases; but if the differences arise from polymorphism or subtle structural differences, the comparison becomes difficult. If structural polymorphism contributes to the difference, then simulated fibre diffraction in Clearer is unable to reproduce this, where the very process of simulated fibre diffraction within Clearer assumes structure monodispersity and a certain degree of crystallinity.

The simulation of X-ray fibre diffraction from models of the crystalline phase and indeed the fibrillar phase is another point raised by this work. Many models of self-assembled LMWG are reported in the literature (Hughes et al., 2011, Smith et al., 2008, Xu et al., 2010) but are not validated by a comparison of simulated data to the originally collected experimental data. The results presented in Chapter 7 illustrate how this step is essential in the validation of proposed models and reveals that models may be logically based upon experimental data but still require further development and refinement.

9.5 Future work

The biophysical and structural techniques presented here are the most rapid way to collect data on the fibrillar phase of amyloid-like self-assembled systems. If the data processing procedures and modelling method with feedback from simulated structural data could be automated, the technique would surpass other structural methods. Despite this XRFD, CD, LD and TEM can only report on specific levels of structure, and where self-assembled systems are hierarchical, this means a range of techniques must still be employed to accurately determine the most representative models of these types of assemblies.

Solid state NMR should be used on all of the systems. This would provide information that could be modelled with the information here. However, it is possible that in the solid state the same questions of comparability to the native fibre phase might be raised. The technique of cryo-EM is developing to a point where there are reports of increasingly higher resolution cryo-EM density maps. If the same could be applied to the systems presented here, these kinds of structural models would surely best represent the fibrillar phase, where samples are inspected in the hydrated state and at concentrations equivalent to those used for assembly.

Where a direct link has not yet been identified between molecular structure and fibrillar structure leading to specific material properties, it is not yet possible to rationally design a monomer with predictable self-assembled material properties. No doubt, there is a determining link because the gels formed respond in a consistent manner to rheological characterisation. Alternatively, the pursuit of tuneable material properties could follow the route of system mixing as briefly described in Chapter 8. It would be very interesting to identify mixtures of dipeptide systems that had differing material properties and investigate the structural basis for this. If this was based on the co-assembly of two dipeptides, the structure could be probed with X-ray fibre diffraction at the short-range. If the basis for this was self-sorting then fibre diffraction at low angles could also give insight. Certainly, where the work presented here is a

thorough characterisation of the structural and biophysical ‘finger prints’ of these systems in their normal assembled and gelled states, any changes found on the macroscopic level could be characterised in the same manner and compared to the molecular, morphological and supramolecular standards established here. It is enticing that the answers to the questions regarding what the molecular basis for differing material properties is remain somewhat unanswered, but a combinatorial approach of new characterisations on different lengths scales with the structural details elucidated here, will surely bring us closer to this goal.

Bibliography

Abramoff, A.M., Magelhaes, P.J., and Ram, S.J. (2004). Image Processing with ImageJ. *Biophotonics International* **11**, 36-42.

Adachi, R., Yamaguchi, K.-i., Yagi, H., Sakurai, K., Naiki, H., and Goto, Y. (2007). Flow-induced alignment of amyloid protofilaments revealed by linear dichroism. *J Biol Chem* **282**, 8978-8983.

Adams, D.J. (2011). Dipeptide and Tripeptide Conjugates as Low-Molecular-Weight Hydrogelators. *Macromol Biosci* **11**, 160-173.

Adams, D.J., Butler, M.F., Frith, W.J., Kirkland, M., Mullen, L., and Sanderson, P. (2009). A new method for maintaining homogeneity during liquid-hydrogel transitions using low molecular weight hydrogelators. *Soft Matter* **5**, 1856-1862.

Adams, D.J., Morris, K., Chen, L., Serpell, L.C., Bacsa, J., and Day, G.M. (2010a). The delicate balance between gelation and crystallisation: structural and computational investigations. *Soft Matter* **6**, 4144-4156.

Adams, D.J., Mullen, L.M., Berta, M., Chen, L., and Frith, W.J. (2010b). Relationship between molecular structure, gelation behaviour and gel properties of Fmoc-dipeptides. *Soft Matter* **6**, 1971-1980.

Aggeli, A., Bell, M., Boden, N., Keen, J.N., McLeish, T.C.B., Nyrkova, I., Radford, S.E., and Semenov, A. (1997). Engineering of peptide ß-sheet nanotapes. *J Mater Chem* **7**, 1135-1145.

Aguzzi, A., and Haass, C. (2003). Games played by rogue proteins in prion disorders and Alzheimer's disease. *Science* **302**, 814-818.

Alexander, L.E. (1969). X-ray Diffraction Methods in Polymer Science. In X-ray Diffraction Methods in Polymer Science (Wiley-interscience).

Amdursky, N., Gazit, E., and Rosenman, G. (2010). Quantum Confinement in Self-Assembled Bioinspired Peptide Hydrogels. *Adv Mater* **22**, 2311-2315.

Amdursky, N., Orbach, R., Gazit, E., and Huppert, D. (2009). Probing the Inner Cavities of Hydrogels by Proton Diffusion. *J Phys Chem C* **113**, 19500-19505.

Amende, M.T., Hariharan, D., and Peppas, N.A. (1995). Factors Influencing Drug and Protein-transport and Release from Ionic Hydrogels. *Reactive Polymers* **25**, 127-137.

Andersen, C.B., Hicks, M.R., Vetri, V., Vandahl, B., Rahbek-Nielsen, H., Thogersen, H., Thogersen, I.B., Enghild, J.J., Serpell, L.C., Rischel, C.*, et al.* (2010). Glucagon fibril polymorphism reflects differences in protofilament backbone structure. *J Mol Biol* **397**, 932-946.

Anfinsen, C.B., Haber, E., Sela, M., and White, F.H. (1961). Kinetics of Formation of Native Ribonuclease During Oxidation of Reduced Polypeptide Chain. *Proc Natl Acad Sci U S A* **47**, 1309-1314.

Astbury, W.T., and Dickinson, S. (1935). The X-ray interpretation of denaturation and the structure of the seed globulins. *Biochem J* **29**, 2351.

Balbirnie, M., Grothe, R., and Eisenberg, D.S. (2001). An amyloid-forming peptide from the yeast prion Sup35 reveals a dehydrated ß-sheet structure for amyloid. *Proc Natl Acad Sci U S A* **98**, 2375-2380.

Baldwin, A.J., Bader, R., Christodoulou, J., MacPhee, C.E., Dobson, C.M., and Barker, P.D. (2006). Cytochrome display on amyloid fibrils. *J Am Chem Soc* **128**, 2162-2163.

Banwell, E.F., Abelardo, E.S., Adams, D.J., Birchall, M.A., Corrigan, A., Donald, A.M., Kirkland, M., Serpell, L.C., Butler, M.F., and Woolfson, D.N. (2009). Rational design and application of responsive α-helical peptide hydrogels. *Nature Materials* **8**, 596-600.

Bartels, T., Choi, J.G., and Selkoe, D.J. (2011). α-Synuclein occurs physiologically as a helically folded tetramer that resists aggregation. *Nature* **477**, 107-123.

Barth, A., and Zscherp, C. (2002). What vibrations tell us about proteins. *Q Rev Biophys* **35**, 369-430.

Bayley, P.M., Haire, L.F., Whyte, S.M., Vasisht, N., Gill, A.C., Verma, C., Dodson, E.J., and Dodson, G.G. (2004). The crystal structure of the globular domain of sheep prion protein. *J Mol Biol* **336**, 1175-1183.

Bertini, I., Gonnelli, L., Luchinat, C., Mao, J., and Nesi, A. (2011). A New Structural Model of Aß(40) Fibrils. *J Am Chem Soc* **133**, 16013-16022.

Bhatia, R., Lin, H., and Lal, R. (2000). Fresh and globular amyloid ß protein (1-42) induces rapid cellular degeneration: evidence for AßP channel-mediated cellular toxicity. *FASEB J* **14**, 1233-1243.

Bian, W., Wang, H., McCullough, I., and Stubbs, G. (2006). WCEN: a computer program for initial processing of fiber diffraction patterns. *J Appl Crystallogr* **39**, 752-756.

Blake, C., and Serpell, L. (1996). Synchrotron X-ray studies suggest that the core of the transthyretin amyloid fibril is a continuous ß-sheet helix. *Structure* **4**, 989-998.

Blow, D.M. (2002). Outline of Crystallography for Biologists (Oxford University Press).

Borovinskiy, A. (2006). DISORDER. http://fibernet.vanderbilt.edu/software/disorder/.

Boyle, A.L., and Woolfson, D.N. (2011). De novo designed peptides for biological applications. *Chem Soc Rev* **40**, 4295-4306.

Broersen, K., Jonckheere, W., Rozenski, J., Vandersteen, A., Pauwels, K., Pastore, A., Rousseau, F., and Schymkowitz, J. (2011). A standardized and biocompatible preparation of aggregate-free amyloid ß peptide for biophysical and biological studies of Alzheimers disease. *Protein Engineering Design & Selection* **24**, 743-750.

Brooks, B.R., Brooks, C.L., III, Mackerell, A.D., Jr., Nilsson, L., Petrella, R.J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S.*, et al.* (2009). CHARMM: The Biomolecular Simulation Program. *J Comput Chem* **30**, 1545-1614.

Bryant, S.J., Nuttelman, C.R., and Anseth, K.S. (2000). Cytocompatibility of UV and visible light photoinitiating systems on cultured NIH/3T3 fibroblasts in vitro. *Journal of Biomaterials Science-Polymer Edition* **11**, 439-457.

Bucak, S., Cenker, C., Nasir, I., Olsson, U., and Zackrisson, M. (2009). Peptide Nanotube Nematic Phase. *Langmuir* **25**, 4262-4265.

Buchan, D.W.A., Ward, S.M., Lobley, A.E., Nugent, T.C.O., Bryson, K., and Jones, D.T. (2010). Protein annotation and modelling servers at University College London. *Nucleic Acids Res* **38**, 563-568.

Bulheller, B.M., Rodger, A., Hicks, M.R., Dafforn, T.R., Serpell, L.C., Marshall, K.E., Bromley, E.H.C., King, P.J.S., Channon, K.J., Woolfson, D.N.*, et al.* (2009). Flow Linear Dichroism of Some Prototypical Proteins. *J Am Chem Soc* **131**, 13305-13314.

Bulheller, B.M., Rodger, A., and Hirst, J.D. (2007). Circular and linear dichroism of proteins. *PCCP* **9**, 2020-2035.

Capaldi, A.P., Kleanthous, C., and Radford, S.E. (2002). Im7 folding mechanism: misfolding on a path to the native state. *Nat Struct Biol* **9**, 209-216.

Chang, X.Q., Jorgensen, A.M.M., Bardrum, P., and Led, J.J. (1997). Solution structures of the R-6 human insulin hexamer. *Biochemistry (Mosc)* **36**, 9409-9422.

Chen, L., Morris, K., Laybourn, A., Elias, D., Hicks, M.R., Rodger, A., Serpell, L., and Adams, D.J. (2010a). Self-Assembly Mechanism for a Naphthalene-Dipeptide Leading to Hydrogelation. *Langmuir* **26**, 5232-5242.

Chen, L., Pont, G., Morris, K., Lotze, G., Squires, A., Serpell, L.C., and Adams, D.J. (2011). Salt-induced hydrogelation of functionalised-dipeptides at high pH. *Chem Commun* **47**, 12071-12073.

Chen, L., Revel, S., Morris, K., and Adams, D.J. (2010b). Energy transfer in self-assembled dipeptide hydrogels. *Chem Commun* **46**, 4267-4269.

Chen, L., Revel, S., Morris, K., Serpell, L.C., and Adams, D.J. (2010c). Effect of Molecular Structure on the Properties of Naphthalene-Dipeptide Hydrogelators. *Langmuir* **26**, 13466-13471.

Cheng, G., Castelletto, V., Moulton, C.M., Newby, G.E., and Hamley, I.W. (2010). Hydrogelation and Self-Assembly of Fmoc-Tripeptides: Unexpected Influence of Sequence on Self-Assembled Fibril Structure, and Hydrogel Modulus and Anisotropy. *Langmuir* **26**, 4990-4998.

Chiti, F., and Dobson, C.M. (2006). Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* **75**, 333-366.

Cody, C.W., Prasher, D.C., Westler, W.M., Prendergast, F.G., and Ward, W.W. (1993). Chemical-structure of the Hexapeptide Chromophore of the Aequorea Green-fluorescent protein. *Biochemistry (Mosc)* **32**, 1212-1218.

Cohen, A.S., and Calkins, E. (1959). Electron Microscopic Observaions on a Fibrous Component in Amyloid of Diverse Origins. *Nature* **183**, 1202-1203.

Colletier, J.-P., Laganowsky, A., Landau, M., Zhao, M., Soriaga, A.B., Goldschmidt, L., Flot, D., Cascio, D., Sawaya, M.R., and Eisenberg, D. (2011). Molecular basis for amyloid-ß polymorphism. *Proc Natl Acad Sci USA* **108**, 16938-16943.

Compton, L.A., and Johnson, W.C. (1986). Analysis of Protein Circular-Dichroism Spectra for Secondary Structure Using a Simple Matrix Multiplication. *Anal Biochem* **155**, 155-167.

Corey, R.B., and Pauling, L. (1953). Fundamental Dimensions of Polypeptide Chains. *Proc R Soc London, Ser B* **141**, 10-20.

Cui, H., Muraoka, T., Cheetham, A.G., and Stupp, S.I. (2009). Self-Assembly of Giant Peptide Nanobelts. *Nano Lett* **9**, 945-951.

Dafforn, T.R., Rajendra, J., Halsall, D.J., Serpell, L.C., and Rodger, A. (2004). Protein fiber linear dichroism for structure determination and kinetics in a low-volume, low-wavelength couette flow cell. *Biophys J* **86**, 404-410.

Dafforn, T.R., and Rodger, A. (2004). Linear dichroism of biomolecules: which way is up? *Curr Opin Struct Biol* **14**, 541-546.

Das, A.K., Collins, R., and Ulijn, R.V. (2008). Exploiting enzymatic (reversed) hydrolysis in directed setf-assembly of peptide nanostructures. *Small* **4**, 279-287.

Dastidar, P. (2008). Supramolecular gelling agents: can they be designed? *Chem Soc Rev* **37**, 2699-2715.

Davidson, W.S., Jonas, A., Clayton, D.F., and George, J.M. (1998). Stabilization of α-synuclein secondary structure upon binding to synthetic membranes. *J Biol Chem* **273**, 9443-9449.

Davis, I.W., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2004). MolProbity: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res* **32**, 615-619.

de Groot, N.S., Pallares, I., Aviles, F.X., Vendrell, J., and Ventura, S. (2005). Prediction of "hot spots" of aggregation in disease-linked polypeptides. *BMC Struct Biol* **5**, 18.

DeLano. (2002). The PyMol Molecular Graphics System. http://www.pymol.org.

Dempsey, J.L., Winkler, J.R., and Gray, H.B. (2010). Mechanism of H(2) Evolution from a Photogenerated Hydridocobaloxime. *J Am Chem Soc* **132**, 16774-16776.

Deshpande, A., Mina, E., Glabe, C., and Busciglio, J. (2006). Different conformations of amyloid-ß induce neurotoxicity by distinct mechanisms in human cortical neurons. *J Neurosci* **26**, 6011-6018.

Diaz-Avalos, R., Long, C., Fontano, E., Balbirnie, M., Grothe, R., Eisenberg, D., and Caspar, D.L.D. (2003). Cross-ß order and diversity in nanocrystals of an amyloid-forming peptide. *J Mol Biol* **330**, 1165-1175.

Dill, K.A., and Chan, H.S. (1997). From Levinthal to pathways to funnels. *Nat Struct Biol* **4**, 10-19.

Dobson, C.M. (2001). The structural basis of protein folding and its links with human disease. *Philos Trans R Soc Lond B Biol Sci* **356**, 133-145.

Dunbrack, R.L., and Cohen, F.E. (1997). Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci* **6**, 1661-1681.

Durell, S.R., Guy, H.R., Arispe, N., Rojas, E., and Pollard, H.B. (1994). Theoretical models of the ion channel structure of amyloid ß-protein. *Biophys J* **67**, 2137-2145.

Eanes, E.D., and Glenner, G.G. (1968). X-ray diffraction studies on amyloid filaments. *J Histochem Cytochem* **16**, 673-677.

Eisenberg, D., Wiltzius, J.J.W., Sievers, S.A., Sawaya, M.R., Cascio, D., Popov, D., and Riekel, C. (2008). Atomic structure of the cross-ß spine of islet amyloid polypeptide (amylin). *Protein Sci* **17**, 1467-1474.

Elisseeff, J., McIntosh, W., Anseth, K., Riley, S., Ragan, P., and Langer, R. (2000). Photoencapsulation of chondrocytes in poly(ethylene oxide)-based semi-interpenetrating networks. *J Biomed Mater Res* **51**, 164-171.

Ellenbogen, E. (1952). Dissociation Constants of Peptides - A Survey of the Effect of Optical Configuration. *J Am Chem Soc* **74**, 5198-5201.

Emerson, E.S., Conlin, M.A., Rosenoff, A.E., Norland, K.S., Rodrigue.H, Chin, D., and Bird, G.R. (1967). Geometrical Structure and Absorption Spectrum of a Cyanine Dye Aggregate. *J Phys Chem* **71**, 2396-2403.

Engler, A.J., Sen, S., Sweeney, H.L., and Discher, D.E. (2006). Matrix elasticity directs stem cell lineage specification. *Cell* **126**, 677-689.

Fandrich, M., and Dobson, C.M. (2002). The behaviour of polyamino acids reveals an inverse side chain effect in amyloid structure formation. *EMBO J* **21**, 5682-5690.

Fasman, G.D. (1996). Circular Dichroism and the Conformational Analysis of Biomolecules (New York and London, Plenum Press).

Ferguson, N., Becker, J., Tidow, H., Tremmel, S., Sharpe, T.D., Krause, G., Flinders, J., Petrovich, M., Berriman, J., Oschkinat, H.*, et al.* (2006). General structural motifs of amyloid protofilaments. *Proc Natl Acad Sci U S A* **103**, 16248-16253.

Fernandez-Escamilla, A.-M., Rousseau, F., Schymkowitz, J., and Serrano, L. (2004). Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat Biotechnol* **22**, 1302-1306.

Ferrone, F. (1999). Analysis of protein aggregation kinetics. *Amyloid, Prions, and Other Protein Aggregates* **309**, 256-274.

Forno, L.S. (1996). Neuropathology of Parkinson's disease. *J Neuropathol Exp Neurol* **55**, 259-272.

Fried, J.R. (2003). Polymer Science & Technology, 2nd edn (Prentice Hall (PTR)).

Galzitskaya, O.V., Garbuzynskiy, S.O., and Lobanov, M.Y. (2006). Prediction of amyloidogenic and disordered regions in protein chains. *PLoS Comp Biol* **2**, 1639-1648.

Gazit, E. (2002). A possible role for pi-stacking in the self-assembly of amyloid fibrils. *FASEB J* **16**, 77-83.

Geddes, A.J., Parker, K.D., Atkins, E.D., and Beighton, E. (1968). "Cross-ß" conformation in proteins. *J Mol Biol* **32**, 343-358.

Ghadiri, M.R., Granja, J.R., Milligan, R.A., McRee, D.E., and Khazanovich, N. (1993). Self-assembling Organic Nanotubes Based on a Cyclic Peptide Architecture. *Nature* **366**, 324-327.

Glenner, G.G., Ein, D., Eanes, E.D., Bladen, H.A., Terry, W., and Page, D.L. (1971). Creation of "amyloid" fibrils from Bence Jones proteins in vitro. *Science* **174**, 712-714.

Glover, J.N.M., and Harrison, S.C. (1995). Crystal-structure of the Heterodimeric BZip Transcriptio Factor C-Fos-C-Jun Bound to DNA. *Nature* **373**, 257-261.

Goedert, M. (2001). α-synuclein and neurodegenerative diseases. *Nat Rev Neurosci* **2**, 492-501.

Goldsbury, C., Goldie, K., Pellaud, J., Seelig, J., Frey, P., Müller, S.A., Kistler, J., Cooper, G.J.S., and Aebi, U. (2000). Amyloid Fibril Formation from Full-Length and Fragments of Amylin. *J Struct Biol* **130**, 352-362.

Goldsbury, C.S., Cooper, G.J.S., Goldie, K.N., Muller, S.A., Saafi, E.L., Gruijters, W.T.M., and Misur, M.P. (1997). Polymorphic fibrillar assembly of human amylin. *J Struct Biol* **119**, 17-27.

Gorzny, M.L., Walton, A.S., and Evans, S.D. (2010). Synthesis of High-Surface-Area Platinum Nanotubes Using a Viral Template. *Adv Funct Mater* **20**, 1295-1300.

Govaerts, C., Wille, H., Prusiner, S.B., and Cohen, F.E. (2004). Evidence for assembly of prions with left-handed ß 3-helices into trimers. *Proc Natl Acad Sci U S A* **101**, 8342-8347.

Gratzer, W.B., Holzwarth, G.M., and Doty, P. (1961). Polarization of the Ultraviolet Absorption Bands in α-helical Polypeptides. *Proc Natl Acad Sci U S A* **47**, 1785-1791.

Greenfield, N.J. (1996). Methods to estimate the conformation of proteins and polypeptides from circular dichroism data. *Anal Biochem* **235**, 1-10.

Grimsley, G.R., Scholtz, J.M., and Pace, C.N. (2009). A summary of the measured pK values of the ionizable groups in folded proteins. *Protein Sci* **18**, 247-251.

Gustavsson, A., Engstrom, U., and Westermark, P. (1991). Normal Transthyretin and Synthetic Transthyretin Fragments Form Amyloid-Like Fibrils Invitro. *Biochem Biophys Res Commun* **175**, 1159-1164.

Hammersley, A.P. (1997). FIT2D: An Introduction and Overview (ESRF Internal Report).

Hamodrakas, S.J., Liappa, C., and Iconomidou, V.A. (2007). Consensus prediction of amyloidogenic determinants in amyloid fibril-forming proteins. *Int J Biol Macromol* **41**, 295-300.

Harada, N., and Nakanishi, K. (1972). Exciton chirality method and its application to configurational and conformational studies of natural products. *Acc Chem Res* **5**, 257-263.

Hardy, J. (2002). Testing times for the "amyloid cascade hypothesis". *Neurobiol Aging* **23**, 1073-1074.

Hardy, J.A., and Higgins, G.A. (1992). Alzheimer's disease: the amyloid cascade hypothesis. *Science* **256**, 184-185.

Harper, J.D., Wong, S.S., Lieber, C.M., and Lansbury, P.T., Jr. (1999). Assembly of Aß amyloid protofibrils: an in vitro model for a possible early event in Alzheimer's disease. *Biochemistry (Mosc)* **38**, 8972-8980.

Hartgerink, J.D., Beniash, E., and Stupp, S.I. (2001). Self-assembly and mineralization of peptide-amphiphile nanofibers. *Science* **294**, 1684-1688.

Hartl, F.U., and Hayer-Hartl, M. (2009). Converging concepts of protein folding in vitro and in vivo. *Nat Struct Mol Biol* **16**, 574-581.

Heise, H., Hoyer, W., Becker, S., Andronesi, O.C., Riedel, D., and Baldus, M. (2005). Molecular-level secondary structure, polymorphism, and dynamics of full-length α-synuclein fibrils studied by solid-state NMR. *Proc Natl Acad Sci U S A* **102**, 15871-15876.

Hicks, M.R., Kowalski, J., and Rodger, A. (2010). LD spectroscopy of natural and synthetic biomaterials. *Chem Soc Rev* **39**, 3380-3393.

Hinsen, K. (2000). The molecular modeling toolkit: A new approach to molecular simulations. *J Comput Chem* **21**, 79-85.

Hiramatsu, H., Goto, Y., Naiki, H., and Kitagawa, T. (2004). Core structure of amyloid fibril proposed from IR-microscope linear dichroism. *J Am Chem Soc* **126**, 3008-3009.

Hofrichter, J., and Eaton, W.A. (1976). Linear Dichroism of Biological Chromophores. *Annu Rev Biophys Bioeng* **5**, 511-560.

Holmes, K.C., Angert, I., Kull, F.J., Jahn, W., and Schroder, R.R. (2003). Electron cryo-microscopy shows how strong binding of myosin to actin releases nucleotide. *Nature* **425**, 423-427.

Holmes, K.C., Popp, D., Gebhard, W., and Kabsch, W. (1990). Atomic model of the actin filament. *Nature* **347**, 44-49.

Houton, K.A., Morris, K.L., Chen, L., Schidtmann, M., Jones, J.T.A., Serpell, L.C., Lloyd, G.O., and Adams, D.J. (2012 - In Press). On Crystal versus Fiber Formation in Dipeptide Hydrogelator Systems. *Langmuir*.

Hughes, M., Xu, H., Frederix, P.W.J.M., Smith, A.M., Hunt, N.T., Tuttle, T., Kinloch, I.A., and Ulijn, R.V. (2011). Biocatalytic self-assembly of 2D peptide-based nanostructures. *Soft Matter* **7**, 10032-10038.

Hwang, W., Zhang, S.G., Kamm, R.D., and Karplus, M. (2004). Kinetic control of dimer structure formation in amyloid fibrillogenesis. *Proc Natl Acad Sci U S A* **101**, 12916-12921.

Ingber, D.E. (2006). Cellular mechanotransduction: putting all the pieces together again. *FASEB J* **20**, 811-827.

Irvine, G.B., El-Agnaf, O.M., Shankar, G.M., and Walsh, D.M. (2008). Protein aggregation in the brain: the molecular basis for Alzheimer's and Parkinson's diseases. *Mol Med* **14**, 451-464.

Ivanova, M.I., Sievers, S.A., Sawaya, M.R., Wall, J.S., and Eisenberg, D. (2009). Molecular basis for insulin fibril assembly. *Proc Natl Acad Sci U S A* **106**, 18990-18995.

Iwata, K., Fujiwara, T., Matsuki, Y., Akutsu, H., Takahashi, S., Naiki, H., and Goto, Y. (2006). 3D structure of amyloid protofilaments of ß2-microglobulin fragment probed by solid-state NMR. *Proc Natl Acad Sci USA* **103**, 18119-18124.

Jahn, T.R., Makin, O.S., Morris, K.L., Marshall, K.E., Tian, P., Sikorski, P., and Serpell, L.C. (2010). The Common Architecture of Cross-ß Amyloid. *J Mol Biol* **395**, 717-727.

Jaroniec, C.P., MacPhee, C.E., Bajaj, V.S., McMahon, M.T., Dobson, C.M., and Griffin, R.G. (2004). High-resolution molecular structure of a peptide in an amyloid fibril determined by magic angle spinning NMR spectroscopy. *Proc Natl Acad Sci U S A* **101**, 711-716.

Jayawarna, V., Ali, M., Jowitt, T.A., Miller, A.E., Saiani, A., Gough, J.E., and Ulijn, R.V. (2006). Nanostructured hydrogels for three-dimensional cell culture through self-assembly of fluorenylmethoxycarbonyl-dipeptides. *Adv Mater* **18**, 611-614.

Jayawarna, V., Richardson, S.M., Hirst, A.R., Hodson, N.W., Saiani, A., Gough, J.E., and Ulijn, R.V. (2009). Introducing chemical functionality in Fmoc-peptide gels for cell culture. *Acta Biomater* **5**, 934-943.

Jia, G.Q., Qiu, S., Li, G.N., Zhou, J., Feng, Z.C., and Li, C. (2009). Alkali-hydrolysis of D-glucono-delta-lactone studied by chiral Raman and circular dichroism spectroscopies. *Science in China Series B-Chemistry* **52**, 552-558.

Jimenez, J.L., Guijarro, J.L., Orlova, E., Zurdo, J., Dobson, C.M., Sunde, M., and Saibil, H.R. (1999). Cryo-electron microscopy structure of an SH3 amyloid fibril and model of the molecular packing. *EMBO J* **18**, 815-821.

Jimenez, J.L., Nettleton, E.J., Bouchard, M., Robinson, C.V., Dobson, C.M., and Saibil, H.R. (2002). The protofilament structure of insulin amyloid fibrils. *Proc Natl Acad Sci U S A* **99**, 9196-9201.

Jimenez, J.L., Tennent, G., Pepys, M., and Saibil, H.R. (2001). Structural diversity of ex vivo amyloid fibrils studied by cryo-electron microscopy. *J Mol Biol* **311**, 241-247.

Johnson, E.K., Adams, D.J., and Cameron, P.J. (2010). Directed Self-Assembly of Dipeptides to Form Ultrathin Hydrogel Membranes. *J Am Chem Soc* **132**, 5130-5136.

Johnson, E.K., Adams, D.J., and Cameron, P.J. (2011). Peptide based low molecular weight gelators. *J Mater Chem* **21**, 2024-2027.

Jones, D.T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* **292**, 195-202.

Joshuator, L., Xu, H.E., Johnston, S.A., and Rees, D.C. (1995). Crystal-structure of a Conserved Protease that Binds DNA - The Bleomycin Hydrolase, GAL6. *Science* **269**, 945-950.

Kabsch, W., Mannherz, H.G., Suck, D., Pai, E.F., and Holmes, K.C. (1990). Atomic structure of the actin:DNase I complex. *Nature* **347**, 37-44.

Kasha, M. (1963). Energy Transfer Mechanisms and Molecular Exciton Model for Molecular Aggregates. *Radiat Res* **20**, 55-70.

Kasha, M., Rawls, H.R., and El-Bayoumi, A. (1965). The Exciton Model in Molecular Spectroscopy. *Pure Appl Chem* **11**, 371-392.

Kelly, J.W. (1996). Alternative conformations of amyloidogenic proteins govern their behavior. *Curr Opin Struct Biol* **6**, 11-17.

Kelly, S.M., Jess, T.J., and Price, N.C. (2005). How to study proteins by circular dichroism. *Biochim Biophys Acta: Proteins Proteomics* **1751**, 119-139.

Kim, Y.M., and Choi, B.-S. (2010). Structure and function of the regulatory HRDC domain from human Bloom syndrome protein. *Nucleic Acids Res* **38**, 7764-7777.

Kirschner, D.A., Abraham, C., and Selkoe, D.J. (1986). X-ray diffraction from intraneuronal paired helical filaments and extraneuronal amyloid fibers in Alzheimer disease indicates cross-ß conformation. *Proc Natl Acad Sci U S A* **83**, 503-507.

Kirschner, D.A., Inouye, H., Duffy, L.K., Sinclair, A., Lind, M., and Selkoe, D.J. (1987). Synthetic peptide homologous to ß protein from Alzheimer disease forms amyloid-like fibrils in vitro. *Proc Natl Acad Sci U S A* **84**, 6953-6957.

Kishimoto, A., Hasegawa, K., Suzuki, H., Taguchi, H., Namba, K., and Yoshida, M. (2004). ß-Helix is a likely core structure of yeast prion Sup35 amyloid fibers. *Biochem Biophys Res Commun* **315**, 739-745.

Kopecek, J. (2009). Hydrogels from Soft Contact Lenses and Implants to Self-Assembled Nanomaterials. *J Polym Sci, Part A: Polym Chem* **47**, 5929-5946.

Kourouklis, G.A., Siomos, K., and Christophorou, L.G. (1982). Vacuum Ultraviolet-Absorption Spectra of Aromatic-Molecules in Solution. *J Mol Spectrosc* **92**, 127-140.

Kyte, J., and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157**, 105-132.

Ladd, M.F.C., and Palmer, R.A. (2003). Structure Determination by X-Ray Crystallography (Kluwer Academic/Plenum Publishers).

Land, M. (2008). Biological optics: Circularly polarised crustaceans. *Curr Biol* **18**, 348-349.

Lee, J., Culyba, E.K., Powers, E.T., and Kelly, J.W. (2011). Amyloid-ß forms fibrils by nucleated conformational conversion of oligomers. *Nat Chem Biol* **7**, 602-609.

Lees, J.G., Miles, A.J., Wien, F., and Wallace, B.A. (2006). A reference database for circular dichroism spectroscopy covering fold and secondary structure space. *Bioinformatics* **22**, 1955-1962.

Leslie, A.G.W. (1992). Recent changes to the MOSFLM package for processing film and image plate data. In Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography.

Levinthal, C. (1968). Are There Pathyways for Protein Folding. *J Chim Phys Phys- Chim Biol* **65**, 44-45.

Lewandowski, J.R., van der Wel, P.C.A., Rigney, M., Grigorieff, N., and Griffin, R.G. (2011). Structural Complexity of a Composite Amyloid Fibril. *J Am Chem Soc* **133**, 14686-14698.

Lide, D.R. (2004). CRC Handbook Chemistry and Physics, 85th edn (CRC Press Inc).

Lin, Y.-H. (2003). Polymer Viscoelasticity - Basics, Molecular Theories and Experiments (World Scientific).

Lopez de la Paz, M., and Serrano, L. (2004). Sequence determinants of amyloid fibril formation. *Proc Natl Acad Sci U S A* **101**, 87-92.

Lorenz, M., Poole, K.J., Popp, D., Rosenbaum, G., and Holmes, K.C. (1995). An atomic model of the unregulated thin filament obtained by X-ray fiber diffraction on oriented actin-tropomyosin gels. *J Mol Biol* **246**, 108-119.

Lorenz, M., Popp, D., and Holmes, K.C. (1993). Refinement of the F-actin model against X-ray fiber diffraction data by the use of a directed mutation algorithm. *J Mol Biol* **234**, 826-836.

Lovell, S.C., Davis, I.W., Adrendall, W.B., de Bakker, P.I.W., Word, J.M., Prisant, M.G., Richardson, J.S., and Richardson, D.C. (2003). Structure validation by C α geometry: phi,psi and C ß deviation. *Proteins-Structure Function and Genetics* **50**, 437-450.

Lu, K., Jacob, J., Thiyagarajan, P., Conticello, V.P., and Lynn, D.G. (2003). Exploiting amyloid fibril lamination for nanotube self-assembly. *J Am Chem Soc* **125**, 6391-6393.

Luhrs, T., Ritter, C., Adrian, M., Riek-Loher, D., Bohrmann, B., Dobeli, H., Schubert, D., and Riek, R. (2005). 3D structure of Alzheimer's amyloid-ß(1-42) fibrils. *Proc Natl Acad Sci U S A* **102**, 17342-17347.

Macdonald, M.E., Ambrose, C.M., Duyao, M.P., Myers, R.H., Lin, C., Srinidhi, L., Barnes, G., Taylor, S.A., James, M., Groot, N.*, et al.* (1993). A Novel Gene Containing a Trinucleotide Repeat That Is Expanded and Unstable on Huntingtons-Disease Chromosomes. *Cell* **72**, 971-983.

Madine, J., Jack, E., Stockley, P.G., Radford, S.E., Serpell, L.C., and Middleton, D.A. (2008). Structural Insights into the Polymorphism of Amyloid-Like Fibrils Formed by Region 20-29 of Amylin Revealed by Solid-State NMR and X-ray Fiber Diffraction. *J Am Chem Soc* **130**, 14990-15001.

Mahler, A., Reches, M., Rechter, M., Cohen, S., and Gazit, E. (2006). Rigid, self-assembled hydrogel composed of a modified aromatic dipeptide. *Adv Mater* **18**, 1365-1370.

Makin, O.S., Atkins, E., Sikorski, P., Johansson, J., and Serpell, L.C. (2005). Molecular basis for amyloid fibril formation and stability. *Proc Natl Acad Sci U S A* **102**, 315-320.

Makin, O.S., and Serpell, L.C. (2004). Structural characterisation of islet amyloid polypeptide fibrils. *J Mol Biol* **335**, 1279-1288.

Makin, O.S., and Serpell, L.C. (2005a). Structures for amyloid fibrils. *FEBS J* **272**, 5950-5961.

Makin, O.S., and Serpell, L.C. (2005b). X-ray diffraction studies of amyloid structure. *Methods in molecular biology (Clifton, NJ)* **299**, 67-80.

Makin, O.S., Sikorski, P., and Serpell, L.C. (2006). Diffraction to study protein and peptide assemblies. *Curr Opin Chem Biol* **10**, 417-422.

Makin, O.S., Sikorski, P., and Serpell, L.C. (2007). CLEARER: a new tool for the analysis of X-ray fibre diffraction patterns and diffraction simulation from atomic structural models. *J Appl Crystallogr* **40**, 966-972.

Mandel, R., and Holzwarth, G. (1973). Ultraviolet circular dichroism of polyproline and oriented collagen. *Biopolymers* **12**, 655-674.

Mao, D., Wachter, E., and Wallace, B.A. (1982). Folding of the Mitochondrial Proton Adenosine-Triphosphatase Proteolipid Channel in Phospholipid-Vesicles. *Biochemistry (Mosc)* **21**, 4960-4968.

Marsh, R.E., Corey, R.B., and Pauling, L. (1955). An Investigation of the Structure of Silk Fibroin. *Biochim Biophys Acta* **16**, 1-34.

Marshall, K.E., Hicks, M.R., Williams, T.L., Hoffmann, S.V., Rodger, A., Dafforn, T.R., and Serpell, L.C. (2010). Characterizing the Assembly of the Sup35 Yeast Prion Fragment, GNNQQNY: Structural Changes Accompany a Fiber-to-Crystal Switch. *Biophys J* **98**, 330-338.

Marshall, K.E., Morris, K.L., Charlton, D., O'Reilly, N., Lewis, L., Walden, H., and Serpell, L.C. (2011). Hydrophobic, Aromatic, and Electrostatic Interactions Play a Central Role in Amyloid Fibril Formation and Stability. *Biochemistry (Mosc)* **50**, 2061-2071.

Marshall, K.E., and Serpell, L.C. (2010). Fibres, crystals and polymorphism: the structural promiscuity of amyloidogenic peptides. *Soft Matter* **6**, 2110-2114.

Mart, R.J., Osborne, R.D., Stevens, M.M., and Ulijn, R.V. (2006). Peptide-based stimuli-responsive biomaterials. *Soft Matter* **2**, 822-835.

Matile, S., Berova, N., Nakanishi, K., Fleischhauer, J., and Woody, R.W. (1996). Structural studies by exciton coupled circular dichroism over a large distance: Porphyrin derivatives of steroids, dimeric steroids, and brevetoxin B. *J Am Chem Soc* **118**, 5198-5206.

Mattson, M.P., Cheng, B., Davis, D., Bryant, K., Lieberburg, I., and Rydel, R.E. (1992). ß-Amyloid peptides destabilize calcium homeostasis and render human cortical neurons vulnerable to excitotoxicity. *J Neurosci* **12**, 376-389.

Maurer-Stroh, S., Debulpaep, M., Kuemmerer, N., de la Paz, M.L., Martins, I.C., Reumers, J., Morris, K.L., Copland, A., Serpell, L., Serrano, L*., et al.* (2010). Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nat Methods* **7**, 237-242.

Maurstad, G., Prass, M., Serpell, L.C., and Sikorski, P. (2009). Dehydration stability of amyloid fibrils studied by AFM. *Eur Biophys J* **38**, 1135-1140.

McDonald, M., Kendalla, A., Tanaka, M., Weissman, J.S., and Stubbs, G. (2008). Enclosed chambers for humidity control and sample containment in fiber diffraction. *J Appl Crystallogr* **41**, 206-209.

McDowell, L.M., and Schaefer, J. (1996). High-resolution NMR of biological solids. *Curr Opin Struct Biol* **6**, 624-629.

McGaughey, G.B., Gagne, M., and Rappe, A.K. (1998). pi-stacking interactions - Alive and well in proteins. *J Biol Chem* **273**, 15458-15463.

Mehta, A.K., Lu, K., Childers, W.S., Liang, Y., Dublin, S.N., Dong, J., Snyder, J.P., Pingali, S.V., Thiyagarajan, P., and Lynn, D.G. (2008). Facial symmetry in protein self-assembly. *J Am Chem Soc* **130**, 9829-9835.

Meinhardt, J., Sachse, C., Hortschansky, P., Grigorieff, N., and Faendrich, M. (2009). Aß(1-40) Fibril Polymorphism Implies Diverse Interaction Patterns in Amyloid Fibrils. *J Mol Biol* **386**, 869-877.

Melki, R., Bousset, L., Belrhali, H., Janin, J., and Morera, S. (2001). Structure of the globular region of the prion protein Ure2 from the yeast Saccharomyces cerevisiae. *Structure* **9**, 39-46.

Meyer, E.A., Castellano, R.K., and Diederich, F. (2003). Interactions with aromatic rings in chemical and biological recognition. *Angew Chem* **42**, 1210-1250.

Miranker, A.D., Calabrese, M.F., Eakin, C.M., and Wang, J.M. (2008). A regulatable switch mediates self-association in an immunoglobulin fold. *Nat Struct Mol Biol* **15**, 965-971.

Miyazawa, T., and Blout, E.R. (1961). Infrared Spectra of Polypeptides in Various Conformations - Amide I and II Bands. *J Am Chem Soc* **83**, 712-719.

Molinspiration. Molinspiration Property Calculator. http://www.molinspiration.com/.

Morgan, C., Colombres, M., Nunez, M.T., and Inestrosa, N.C. (2004). Structure and function of amyloid in Alzheimer's disease. *Prog Neurobiol* **74**, 323-349.

Morris, K., and Serpell, L. (2010). From natural to designer self-assembling biopolymers, the structural characterisation of fibrous proteins & peptides using fibre diffraction. *Chem Soc Rev* **39**, 3445-3453.

Morris, K.L., and Serpell, L. (in press). From molecular to supramolecular amyloid structures: contributions from fiber diffraction and electron microscopy. In Amyloid Fibrils and Prefibrillar Aggregates Molecular and Biological Properties, D.E. Otzen, ed. (Aarhus, Denmark, Viley-VCH).

Morris, K.L., and Serpell, L.C. (2012). X-ray fibre diffraction studies of amyloid fibrils. In Amyloid Proteins: Methods and Protocols (Totowa, NJ, Humana Press Inc).

Mukamel, S., Tretiak, S., Wagersreiter, T., and Chernyak, V. (1997). Electronic coherence and collective optical excitations of conjugated molecules. *Science* **277**, 781-787.

Muro-Small, M.L., Chen, J., and McNeil, A.J. (2011). Dissolution Parameters Reveal Role of Structure and Solvent in Molecular Gelation. *Langmuir* **27**, 13248-13253.

Murphy, R.M. (2007). Kinetics of amyloid formation and membrane interaction with amyloidogenic proteins. *Biochim Biophys Acta* **1768**, 1923-1934.

Naskar, J., Palui, G., and Banerjee, A. (2009). Tetrapeptide-Based Hydrogels: for Encapsulation and Slow Release of an Anticancer Drug at Physiological pH. *J Phys Chem B* **113**, 11787-11792.

Nelson, R., Sawaya, M.R., Balbirnie, M., Madsen, A.O., Riekel, C., Grothe, R., and Eisenberg, D. (2005). Structure of the cross-ß spine of amyloid-like fibrils. *Nature* **435**, 773-778.

Neri, C., Holbert, S., Denghien, I., Kiechle, T., Rosenblatt, A., Wellington, C., Hayden, M.R., Margolis, R.L., Ross, C.A., Dausset, J.*, et al.* (2001). The Gln-Ala repeat transcriptional activator CA150 interacts with huntingtin: Neuropathologic and genetic evidence for a role in Huntington's disease pathogenesis. *Proc Natl Acad Sci U S A* **98**, 1811-1816.

Norden, B. (1977). Linear and circular dichroism of polymeric pseudoisocyanine. *J Phys Chem* **81**, 151-159.

Nordh, J., Deinum, J., and Norden, B. (1986). Flow Orientation of Brain Microtubules Studied by Linear Dichroism. *European Biophysics Journal with Biophysics Letters* **14**, 113-122.

Oancea, S., Formaggio, F., Campestrini, S., Broxterman, Q.B., Kaptein, B., and Toniolo, C. (2003). Distance dependency of exciton coupled circular dichroism using turn and helical peptide spacers. *Biopolymers* **72**, 105-115.

Oda, T., Iwasa, M., Aihara, T., Maeda, Y., and Narita, A. (2009). The nature of the globular- to fibrous-actin transition. *Nature* **457**, 441-445.

Pandya, M.J., Spooner, G.M., Sunde, M., Thorpe, J.R., Rodger, A., and Woolfson, D.N. (2000). Sticky-end assembly of a designed peptide fiber provides insight into protein fibrillogenesis. *Biochemistry (Mosc)* **39**, 8728-8734.

Papapostolou, D., Smith, A.M., Atkins, E.D., Oliver, S.J., Ryadnov, M.G., Serpell, L.C., and Woolfson, D.N. (2007). Engineering nanoscale order into a designed protein fiber. *Proc Natl Acad Sci U S A* **104**, 10853-10858.

Pappenheimer, J.R. (1953). Passage of Molecules Through Capillary Walls. *Physiol Rev* **33**, 387-423.

Paravastu, A.K., Leapman, R.D., Yau, W.M., and Tycko, R. (2008). Molecular structural basis for polymorphism in Alzheimer's ß-amyloid fibrils. *Proc Natl Acad Sci U S A* **105**, 18349-18354.

Paravastu, A.K., Petkova, A.T., and Tycko, R. (2006). Polymorphic fibril formation by residues 10-40 of the Alzheimer's ß-amyloid peptide. *Biophys J* **90**, 4618-4629.

Pauling, L., and Corey, R.B. (1953a). Compoung Helical Configurations of Polypeptide Chains - Structure of Proteins of the α-Keratin type. *Nature* **171**, 59-61.

Pauling, L., and Corey, R.B. (1953b). Stable Configurations of Polypeptide Chains. *Proceedings of the Royal Society of London Series B-Biological Sciences* **141**, 21-33.

Pawar, A.P., DuBay, K.F., Zurdo, J., Chiti, F., Vendruscolo, M., and Dobson, C.M. (2005). Prediction of "aggregation-prone" and "aggregation-susceptible" regions in proteins associated with neurodegenerative diseases. *J Mol Biol* **350**, 379-392.

Pelton, J.T., and McLean, L.R. (2000). Spectroscopic Methods for Analysis of Protein Secondary Structure. *Anal Biochem* **277**, 167-176.

Peppas, N.A., Slaughter, B.V., Khurshid, S.S., Fisher, O.Z., and Khademhosseini, A. (2009). Hydrogels in Regenerative Medicine. *Adv Mater* **21**, 3307-3329.

Perutz, M.F., Finch, J.T., Berriman, J., and Lesk, A. (2002). Amyloid fibers are water-filled nanotubes. *Proc Natl Acad Sci U S A* **99**, 5591-5595.

Petkova, A.T., Ishii, Y., Balbach, J.J., Antzutkin, O.N., Leapman, R.D., Delaglio, F., and Tycko, R. (2002). A structural model for Alzheimer's ß-amyloid fibrils based on experimental constraints from solid state NMR. *Proc Natl Acad Sci U S A* **99**, 16742-16747.

Petkova, A.T., Leapman, R.D., Guo, Z., Yau, W.M., Mattson, M.P., and Tycko, R. (2005). Self-propagating, molecular-level polymorphism in Alzheimer's ß-amyloid fibrils. *Science* **307**, 262-265.

Petkova, A.T., Yau, W.M., and Tycko, R. (2006). Experimental constraints on quaternary structure in Alzheimer's ß-amyloid fibrils. *Biochemistry (Mosc)* **45**, 498-512.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF chimera - A visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612.

Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kale, L., and Schulten, K. (2005). Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781-1802.

Picone, D., Crescenzi, O., Tomaselli, S., Guerrini, R., Salvadori, S., D'Ursi, A.M., and Temussi, P.A. (2002). Solution structure of the Alzheimer amyloid ß-peptide (1-42) in an apolar microenvironment - Similarity with a virus fusion domain. *Eur J Biochem* **269**, 5642-5648.

Pocker, Y., and Green, E. (1973). Hydrolysis of D-Glucono-Delta-Lactone .1. General Acid-Base Catalysis, Solvent Deuterium-Isotope Effects, and Transition-State Characterization. *J Am Chem Soc* **95**, 113-119.

Qiu, Z.J., Yu, H.T., Li, J.B., Wang, Y., and Zhang, Y. (2009). Spiropyran-linked dipeptide forms supramolecular hydrogel with dual responses to light and to ligand-receptor interaction. *Chem Commun*, 3342-3344.

Radford, S.E., Dobson, C.M., and Evans, P.A. (1992). The Folding of Hen Lysoyme Involves Partially Structured Intermediates and Multiple Pathways. *Nature* **358**, 302-307.

Reches, M., and Gazit, E. (2006). Controlled patterning of aligned self-assembled peptide nanotubes. *Nature Nanotechnology* **1**, 195-200.

Resende, R., Ferreiro, E., Pereira, C., and Resende de Oliveira, C. (2008). Neurotoxic effect of oligomeric and fibrillar species of amyloid-ß peptide 1-42: involvement of endoplasmic reticulum calcium release in oligomer-induced cell death. *Neuroscience* **155**, 725-737.

Reumers, J., Maurer-Stroh, S., Schymkowitz, J., and Rousseau, F. (2009). Protein Sequences Encode Safeguards Against Aggregation. *Hum Mutat* **30**, 431-437.

Rhodes, W. (1961). Hypochromism and Other Spectral Properties of Helical Polynucleotides. *J Am Chem Soc* **83**, 3609-3617.

Robert W, W. (1993). The circular dichroism of oriented ß-sheets: Theoretical predictions. *Tetrahedron: Asymmetry* **4**, 529-544.

Rodger, A. (1993). Linear dichroism. In Methods Enzymol, F.R. James, and L.V. Bert, eds. (Academic Press), pp. 232-258.

Rodger, A., and Norden, B. (1997). Circular Chroism & Linear Dichroism (OUP Oxford).

Rodger, A., Rajendra, J., Marrington, R., Ardhammar, M., Norden, B., Hirst, J.D., Gilbert, A.T.B., Dafforn, T.R., Halsall, D.J., Woolhead, C.A.*, et al.* (2002). Flow oriented linear dichroism to probe protein orientation in membrane environments. *PCCP* **4**, 4051-4057.

Rughani, R.V., Salick, D.A., Lamm, M.S., Yucel, T., Pochan, D.J., and Schneider, J.P. (2009). Folding, self-assembly, and bulk material properties of a de novo designed three-stranded ß-sheet hydrogel. *Biomacromolecules* **10**, 1295-1304.

Rughani, R.V., and Schneider, J.P. (2008). Molecular design of ß-hairpin peptides for material construction. *MRS Bull* **33**, 530-535.

Sachse, C., Fandrich, M., and Grigorieff, N. (2008). Paired ß-sheet structure of an Aß(1-40) amyloid fibril revealed by electron microscopy. *Proc Natl Acad Sci U S A* **105**, 7462-7466.

Sachse, C., Xu, C., Wieligmann, K., Diekmann, S., Grigorieff, N., and Fandrich, M. (2006). Quaternary structure of a mature amyloid fibril from Alzheimer's Aß(1-40) peptide. *J Mol Biol* **362**, 347-354.

Saibil, H.R. (2007). How to Read Papers on Three-Dimensional Structure Determination by Electron Microscopy. In Evaluating Techniques in Biochemical Research, D Zuk, ed (Cambridge, MA: Cell Press).

Saiki, M., Konakahara, T., and Morii, H. (2006). Interaction-based evaluation of the propensity for amyloid formation with cross-ß structure. *Biochem Biophys Res Commun* **343**, 1262-1271.

Salonen, L.M., Ellermann, M., and Diederich, F. (2011). Aromatic Rings in Chemical and Biological Recognition: Energetics and Structures. *Angew Chem* **50**, 4808-4842.

Sauer-Eriksson, A.E., Eneqvist, T., Andersson, K., Olofsson, A., and Lundgren, E. (2000). The ß-slip: A novel concept in transthyretin amyloidosis. *Mol Cell* **6**, 1207-1218.

Sawaya, M.R., Sambashivan, S., Nelson, R., Ivanova, M.I., Sievers, S.A., Apostol, M.I., Thompson, M.J., Balbirnie, M., Wiltzius, J.J., McFarlane, H.T., *et al.* (2007). Atomic structures of amyloid cross-ß spines reveal varied steric zippers. *Nature* **447**, 453-457.

Schneider, J.P., Pochan, D.J., Ozbas, B., Rajagopal, K., Pakstis, L., and Kretsinger, J. (2002). Responsive hydrogels from the intramolecular folding and self-assembly of a designed peptide. *J Am Chem Soc* **124**, 15030-15037.

Scholes, G.D., Ghiggino, K.P., Oliver, A.M., and Paddonrow, M.N. (1993). Through-space and Through-bond Effects on Exciton Interactions in Rigidly Linked Dinaphthyl Molecules. *J Am Chem Soc* **115**, 4345-4349.

Scholes, G.D., Turner, G.O., Ghiggino, K.P., Paddon-Row, M.N., Piet, J.J., Schuddeboom, W., and Warman, J.M. (1998). Electronic interactions in rigidly linked naphthalene dimers. *Chem Phys Lett* **292**, 601-606.

Seilheimer, B., Bohrmann, B., Bondolfi, L., Muller, F., Stuber, D., and Dobeli, H. (1997). The toxicity of the Alzheimer's ß-amyloid peptide correlates with a distinct fiber morphology. *J Struct Biol* **119**, 59-71.

Selkoe, D.J. (1990). Deciphering Alzheimer's Disease - The Amyloid Precursor Protein Yields New Clues. *Science* **248**, 1058-1060.

Serpell, L.C. (2000). Alzheimer's amyloid fibrils: structure and assembly. *Biochim Biophys Acta* **1502**, 16-30.

Serpell, L.C., Berriman, J., Jakes, R., Goedert, M., and Crowther, R.A. (2000a). Fiber diffraction of synthetic α-synuclein filaments shows amyloid-like cross-ß conformation. *Proc Natl Acad Sci U S A* **97**, 4897-4902.

Serpell, L.C., and Smith, J.M. (2000). Direct visualisation of the ß-sheet structure of synthetic Alzheimer's amyloid. *J Mol Biol* **299**, 225-231.

Serpell, L.C., Sunde, M., Benson, M.D., Tennent, G.A., Pepys, M.B., and Fraser, P.E. (2000b). The protofilament substructure of amyloid fibrils. *J Mol Biol* **300**, 1033-1039.

Serpell, L.C., Sunde, M., Fraser, P.E., Luther, P.K., Morris, E.P., Sangren, O., Lundgren, E., and Blake, C.C.F. (1995). Examination of the Structure of the Transthyretin Amyloid Fibril by Image-Reconstruction from Electron-Micrographs. *J Mol Biol* **254**, 113-118.

Shao, H., Gao, M., Kim, S.H., Jaroniec, C.P., and Parquette, J.R. (2011). Aqueous Self-Assembly of L-Lysine-Based Amphiphiles into 1D n-Type Nanotubes. *Chemistry-a European Journal* **17**, 12882-12885.

Shao, H., Jao, S., Ma, K., and Zagorski, M.G. (1999). Solution structures of micelle-bound amyloid ß-(1-40) and ß-(1-42) peptides of Alzheimer's disease. *J Mol Biol* **285**, 755-773.

Shiraham, T., and Cohen, A.S. (1965). Structure of Amyloid Fibrils after Negative Staining and High-resolution Electron Microscopy. *Nature* **206**, 737-738.

Sikorski, P., and Atkins, E. (2005). New model for crystalline polyglutamine assemblies and their connection with amyloid fibrils. *Biomacromolecules* **6**, 425-432.

Sikorski, P., Atkins, E.D.T., and Serpell, L.C. (2003). Structure and texture of fibrous crystals formed by Alzheimer's Aß(11-25) peptide fragment. *Structure* **11**, 915-926.

Simonyi, M., Bikádi, Z., Zsila, F., and Deli, J. (2003). Supramolecular exciton chirality of carotenoid aggregates. *Chirality* **15**, 680-698.

Singh, R.A., and Thakur, S.N. (1981). Electronic-Transition Moments in the Spectra of Substituted Naphthalenes. *Journal of Crystal and Molecular Structure* **11**, 197-201.

Smith, A.M., Williams, R.J., Tang, C., Coppo, P., Collins, R.F., Turner, M.L., Saiani, A., and Ulijn, R.V. (2008). Fmoc-Diphenylalanine self assembles to a hydrogel via a novel architecture based on π-π interlocked ß-sheets. *Adv Mater* **20**, 37-41.

Smith, K.J., Reid, S.W., Harlos, K., McMichael, A.J., Stuart, D.I., Bell, J.I., and Jones, E.Y. (1996). Bound Water Structure and Polymorphic Amino Acids Act Together to Allow the Binding of Different Peptides to MHC Class I HLA-B53. *Immunity* **4**, 215-228.

Sokolov, Y., Kozak, J.A., Kayed, R., Chanturiya, A., Glabe, C., and Hall, J.E. (2006). Soluble amyloid oligomers increase bilayer conductance by altering dielectric structure. *J Gen Physiol* **128**, 637-647.

Spillantini, M.G., Schmidt, M.L., Lee, V.M.Y., Trojanowski, J.Q., Jakes, R., and Goedert, M. (1997). α-Synuclein in Lewy bodies. *Nature* **388**, 839-840.

Squires, A.M., Devlin, G.L., Gras, S.L., Tickler, A.K., MacPhee, C.E., and Dobson, C.M. (2006). X-ray scattering study of the effect of hydration on the cross-ß structure of amyloid fibrils. *J Am Chem Soc* **128**, 11738-11739.

Sreerama, N., and Woody, R.W. (2004). Computation and Analysis of Protein Circular Dichroism Spectra. In Methods Enzymol, B. Ludwig, and L.J. Michael, eds. (Academic Press), pp. 318-351.

Steinbacher, S., Seckler, R., Miller, S., Steipe, B., Huber, R., and Reinemer, P. (1994). Crystal-Structure of P22 Tailspike Protein - Interdigitated Subunits in a Thermostable Trimer. *Science* **265**, 383-386.

Steinmetz, M.O., Gattin, Z., Verel, R., Ciani, B., Stromer, T., Green, J.M., Tittmann, P., Schulze-Briese, C., Gross, H., van Gunsteren, W.F*., et al.* (2008). Atomic models of de novo designed cc ß-Met amyloid-like fibrils. *J Mol Biol* **376**, 898-912.

Steven, A.C., Kajava, A.V., Baxa, U., and Wickner, R.B. (2004). A model for Ure2p prion filaments and other amyloids: The parallel superpleated ß-structure. *Proc Natl Acad Sci U S A* **101**, 7885-7890.

Sticht, H., Bayer, P., Willbold, D., Dames, S., Hilbich, C., Beyreuther, K., Frank, R.W., and Rosch, P. (1995). Structure of Amyloid A4-(1-40)-Peptide of Alzheimers-Disease. *Eur J Biochem* **233**, 293-298.

Straub, F.B. (1942). Actin. *Studies, University of Szeged* **II**, 1-15.

Sutton, S., Campbell, N.L., Cooper, A.I., Kirkland, M., Frith, W.J., and Adams, D.J. (2009). Controlled Release from Modified Amino Acid Hydrogels Governed by Molecular Size or Network Dynamics. *Langmuir* **25**, 10285-10291.

Tang, C., Smith, A.M., Collins, R.F., Ulijn, R.V., and Saiani, A. (2009). Fmoc-Diphenylalanine Self-Assembly Mechanism Induces Apparent pKa Shifts. *Langmuir* **25**, 9447-9453.

Tartaglia, G.G., Cavalli, A., Pellarin, R., and Caflisch, A. (2005). Prediction of aggregation rate and aggregation-prone segments in polypeptide sequences. *Protein Sci* **14**, 2723-2734.

Tattum, M.H., Cohen-Krausz, S., Khalili-Shirazi, A., Jackson, G.S., Orlova, E.V., Collinge, J., Clarke, A.R., and Saibil, H.R. (2006). Elongated Oligomers Assemble into Mammalian PrP Amyloid Fibrils. *J Mol Biol* **357**, 975-985.

Thompson, M.J., Sievers, S.A., Karanicolas, J., Ivanova, M.I., Baker, D., and Eisenberg, D. (2006). The 3D profile method for identifying fibril-forming segments of proteins. *Proc Natl Acad Sci U S A* **103**, 4074-4078.

Toledano, S., Williams, R.J., Jayawarna, V., and Ulijn, R.V. (2006). Enzyme-triggered self-assembly of peptide hydrogels via reversed hydrolysis. *J Am Chem Soc* **128**, 1070-1071.

Townend, R., Kumosinski, T.F., Timasheff, S.N., Fasman, G.D., and Davidson, B. (1966). The circular dichroism of the ß structure of poly-l-lysine. *Biochem Biophys Res Commun* **23**, 163-169.

Trinh, C.H., Smith, D.P., Kalverda, A.P., Phillips, S.E.V., and Radford, S.E. (2002). Crystal structure of monomeric human ß-2-microglobulin reveals clues to its amyloidogenic properties. *Proc Natl Acad Sci U S A* **99**, 9771-9776.

Tycko, R. (2000). Solid-state NMR as a probe of amyloid fibril structure. *Curr Opin Chem Biol* **4**, 500-506.

Tycko, R., Luca, S., Yau, W.M., and Leapman, R. (2007). Peptide conformation and supramolecular organization in amylin fibrils: Constraints from solid-state NMR. *Biochemistry (Mosc)* **46**, 13505-13522.

Ulijn, R.V., and Smith, A.M. (2008). Designing peptide based nanomaterials. *Chem Soc Rev* **37**, 664-675.

Uri, S. (2008). International Tables for Crystallography, Volume B (Springer).

Urry, D.W., Peng, S.Q., Parker, T.M., Gowda, D.C., and Harris, R.D. (1993). Relative Significance of Electrostatic-Induced and Hydrophobic-Induced Pk(a) Shifts in a Model Protein - the Aspartic-Acid Residue. *Angew Chem Int Ed Engl* **32**, 1440-1442.

Valery, C., Artzner, F., and Paternostre, M. (2011). Peptide nanotubes: molecular organisations, self-assembly mechanisms and applications. *Soft Matter* **7**, 9583-9594.

Valery, C., Paternostre, M., Robert, B., Gulik-Krzywicki, T., Narayanan, T., Dedieu, J.C., Keller, G., Torres, M.L., Cherif-Cheikh, R., Calvo, P.*, et al.* (2003). Biomimetic organization: Octapeptide self-assembly into nanotubes of viral capsid-like dimension. *Proc Natl Acad Sci U S A* **100**, 10258-10262.

van der Wel, P.C., Lewandowski, J.R., and Griffin, R.G. (2007). Solid-state NMR study of amyloid nanocrystals and fibrils formed by the peptide GNNQQNY from yeast prion protein Sup35p. *J Am Chem Soc* **129**, 5117-5130.

van Esch, J.H. (2009). We Can Design Molecular Gelators, But Do We Understand Them? *Langmuir* **25**, 8392-8394.

Veerman, C., Rajagopal, K., Palla, C.S., Pochan, D.J., Schneider, J.P., and Furst, E.M. (2006). Gelation kinetics of ß-hairpin peptide hydrogel networks. *Macromolecules* **39**, 6608-6614.

Vegners, R., Shestakova, I., Kalvinsh, I., Ezzell, R.M., and Janmey, P.A. (1995). Use of a gel-forming dipeptide derivative as a carrier for antigen presentation. *Journal of peptide science : an official publication of the European Peptide Society* **1**, 371-378.

Vibert, P., Craig, R., and Lehman, W. (1997). Steric-model for activation of muscle thin filaments. *J Mol Biol* **266**, 8-14.

Vilar, M., Chou, H.T., Luhrs, T., Maji, S.K., Riek-Loher, D., Verel, R., Manning, G., Stahlberg, H., and Riek, R. (2008). The fold of α-synuclein fibrils. *Proc Natl Acad Sci U S A* **105**, 8637-8642.

Wang, J.M., Cieplak, P., and Kollman, P.A. (2000). How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J Comput Chem* **21**, 1049-1074.

Wang, Q., Yang, Z., Gao, Y., Ge, W., Wang, L., and Xu, B. (2008). Enzymatic hydrogelation to immobilize an enzyme for high activity and stability. *Soft Matter* **4**, 550-553.

Wang, Q., Yang, Z., Wang, L., Ma, M., and Xu, B. (2007). Molecular hydrogel-immobilized enzymes exhibit superactivity and high stability in organic solvents. *Chem Commun*, 1032-1034.

Wasmer, C., Lange, A., Van Melckebeke, H., Siemer, A.B., Riek, R., and Meier, B.H. (2008). Amyloid fibrils of the HET-s(218-289) prion form a ß solenoid with a triangular hydrophobic core. *Science* **319**, 1523-1526.

Weinberg, R.B. (1988). Exposure and Electronic Interaction of Tyrosine and Tryptophan Residues in Human Apolipoprotein A-IV. *Biochemistry (Mosc)* **27**, 1515-1521.

Weinreb, P.H., Zhen, W.G., Poon, A.W., Conway, K.A., and Lansbury, P.T. (1996). NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. *Biochemistry (Mosc)* **35**, 13709-13715.

White, H.E., Hodgkinson, J.L., Jahn, T.R., Cohen-Krausz, S., Gosal, W.S., Mueller, S., Orlova, E.V., Radford, S.E., and Saibil, H.R. (2009). Globular Tetramers of ß2-Microglobulin Assemble into Elaborate Amyloid Fibrils. *J Mol Biol* **389**, 48-57.

Whitesides, G.M., Mathias, J.P., and Seto, C.T. (1991). Molecular Self-Assembly and Nanochemistry - a Chemical Strategy for the Synthesis of Nanostructures. *Science* **254**, 1312-1319.

Whitmore, L., and Wallace, B.A. (2004). DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res* **32**, 668-673.

Wille, H., Bian, W., McDonald, M., Kendall, A., Colby, D.W., Bloch, L., Ollesch, J., Borovinskiy, A.L., Cohen, F.E., Prusiner, S.B.*, et al.* (2009). Natural and synthetic prion structure from X-ray fiber diffraction. *Proc Natl Acad Sci U S A* **106**, 16990-16995.

Williams, T.L., and Serpell, L.C. (2011). Membrane and surface interactions of Alzheimer's Aß peptide - insights into the mechanism of cytotoxicity. *FEBS J* **278**, 3905-3917.

Wiltzius, J.J.W., Sievers, S.A., Sawaya, M.R., and Eisenberg, D. (2009). Atomic structures of IAPP (amylin) fusions suggest a mechanism for fibrillation and the role of insulin in the process. *Protein Sci* **18**, 1521-1530.

Woody, R.W. (1996). Theory of Circular Dichroism of Proteins (New York and London, plenum Press).

Woody, R.W. (2007). Aromatic Side-Chain Contributions to Protein Circular Dichroism (Nova Science Publishers, Inc).

Würthner, F., Kaiser, T.E., and Saha-Möller, C.R. (2011). J-Aggregates: From Serendipitous Discovery to Supramolecular Engineering of Functional Dye Materials. *Angew Chem Int Ed* **50**, 3376-3410.

Xie, Z., Zhang, A., Ye, L., Wang, X., and Feng, Z.-g. (2009). Shear-assisted hydrogels based on self-assembly of cyclic dipeptide derivatives. *J Mater Chem* **19**, 6100-6102.

Xu, B., Yang, Z.M., Liang, G.L., Ma, M.L., and Gao, Y. (2007). Conjugates of naphthalene and dipeptides produce molecular hydrogelators with high efficiency of hydrogelation and superhelical nanofibers. *J Mater Chem* **17**, 850-854.

Xu, H.X., Das, A.K., Horie, M., Shaik, M.S., Smith, A.M., Luo, Y., Lu, X.F., Collins, R., Liem, S.Y., Song, A.M*., et al.* (2010). An investigation of the conductivity of peptide nanotube networks prepared by enzyme-triggered self-assembly. *Nanoscale* **2**, 960-966.

Xue, W.-F., Hellewell, A.L., Gosal, W.S., Homans, S.W., Hewitt, E.W., and Radford, S.E. (2009). Fibril fragmentation enhances amyloid cytotoxicity. *J Biol Chem* **284**, 34272-34282.

Xue, W.-F., Hellewell, A.L., Hewitt, E.W., and Radford, S.E. (2010). Fibril fragmentation in amyloid assembly and cytotoxicity When size matters. *Prion* **4**, 20-25.

Xue, W.-F., Homans, S.W., and Radford, S.E. (2008). Systematic analysis of nucleation-dependent polymerization reveals new insights into the mechanism of amyloid self-assembly. *Proc Natl Acad Sci USA* **105**, 8926-8931.

Yamaguchi, K.-i., Takahashi, S., Kawai, T., Naiki, H., and Goto, Y. (2005). Seeding-dependent Propagation and Maturation of Amyloid Fibril Conformation. *J Mol Biol* **352**, 952-960.

Yang, F., Moss, L.G., and Phillips, G.N. (1996). The molecular structure of green fluorescent protein. *Nat Biotechnol* **14**, 1246-1251.

Yang, Z., Gu, H., Zhang, Y., Wang, L., and Xu, B. (2004). Small molecule hydrogels based on a class of antiinflammatory agents. *Chem Commun*, 208-209.

Yang, Z., and Xu, B. (2004). A simple visual assay based on small molecule hydrogels for detecting inhibitors of enzymes. *Chem Commun*, 2424-2425.

Yang, Z., Xu, K., Wang, L., Gu, H., Wei, H., Zhang, M., and Xu, B. (2005). Self-assembly of small molecules affords multifunctional supramolecular hydrogels for topically treating simulated uranium wounds. *Chem Commun*, 4414-4416.

Yatin, S.M., Varadarajan, S., Link, C.D., and Butterfield, D.A. (1999). In vitro and in vivo oxidative stress associated with Alzheimer's amyloid ß-peptide (1-42). *Neurobiol Aging* **20**, 325-330; discussion 339-342.

Zbigniew K, W. (2002). III - Rheology. In Processes of Fiber Formation (Oxford, Elsevier Science Ltd), pp. 44-99.

Zeltinger, J., Sherwood, J.K., Graham, D.A., Mueller, R., and Griffith, L.G. (2001). Effect of pore size and void fraction on cellular adhesion, proliferation, and matrix deposition. *Tissue Eng* **7**, 557-572.

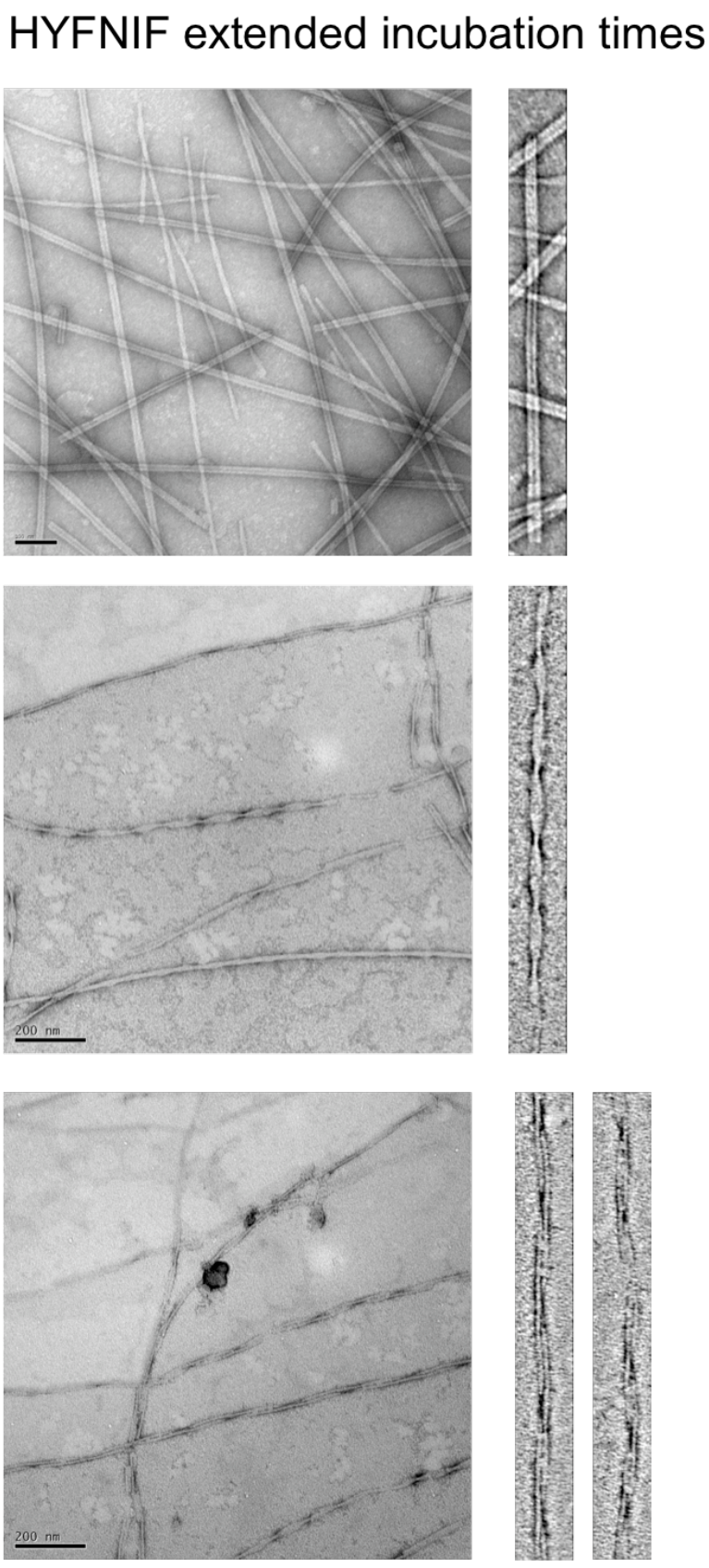
Zhang, Y., Gu, H.W., Yang, Z.M., and Xu, B. (2003). Supramolecular hydrogels respond to ligand-receptor interaction. *J Am Chem Soc* **125**, 13680-13681.

Zhang, Y., Yang, Z.M., Yuan, F., Gu, H.W., Gao, P., and Xu, B. (2004). Molecular recognition remolds the self-assembly of hydrogelators and increases the elasticity of the hydrogel by 10(6)-fold. *J Am Chem Soc* **126**, 15028-15029.

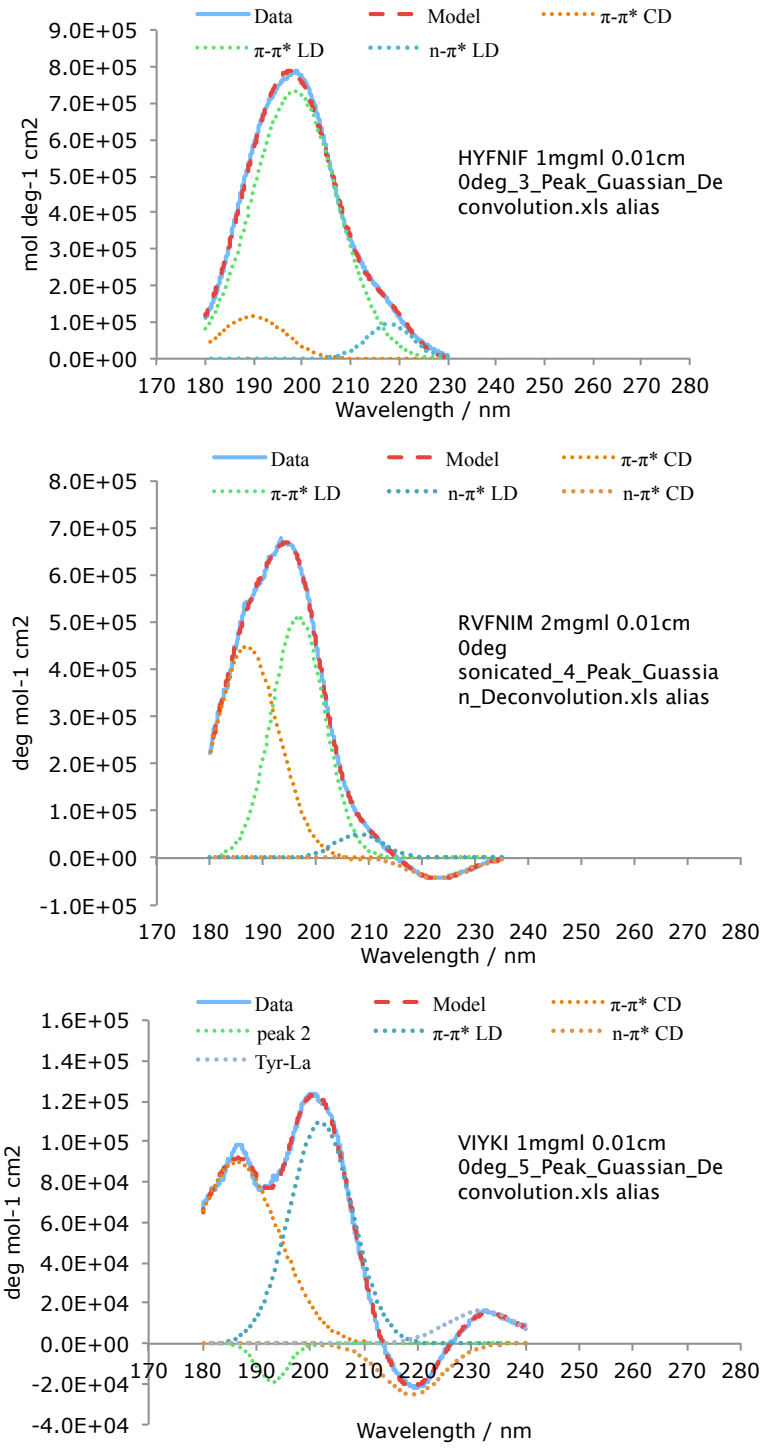
Zhou, M., Smith, A.M., Das, A.K., Hodson, N.W., Collins, R.F., Ulijn, R.V., and Gough, J.E. (2009). Self-assembled peptide-based hydrogels as scaffolds for anchorage-dependent cells. *Biomaterials* **30**, 2523-2530.

Zibaee, S., Makin, O.S., Goedert, M., and Serpell, L.C. (2007). A simple algorithm locates ß-strands in the amyloid fibril core of α-synuclein, Aß, and tau using the amino acid sequence alone. *Protein Sci* **16**, 906-918.

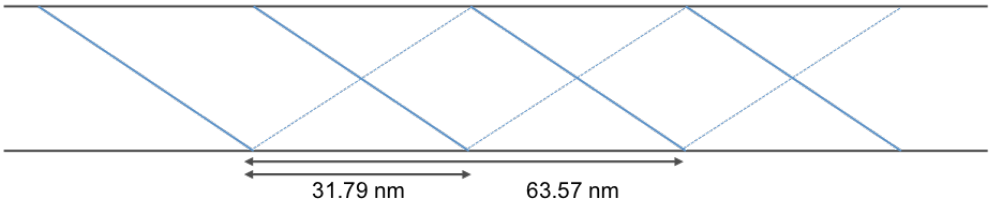
Appendix I

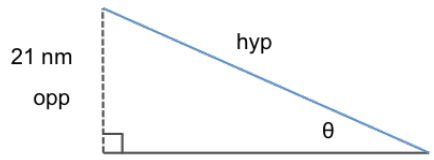


Waltz peptide, 0 deg pre sonication LD artefact deconvolution



VIYKI TEM/LD helicity analysis





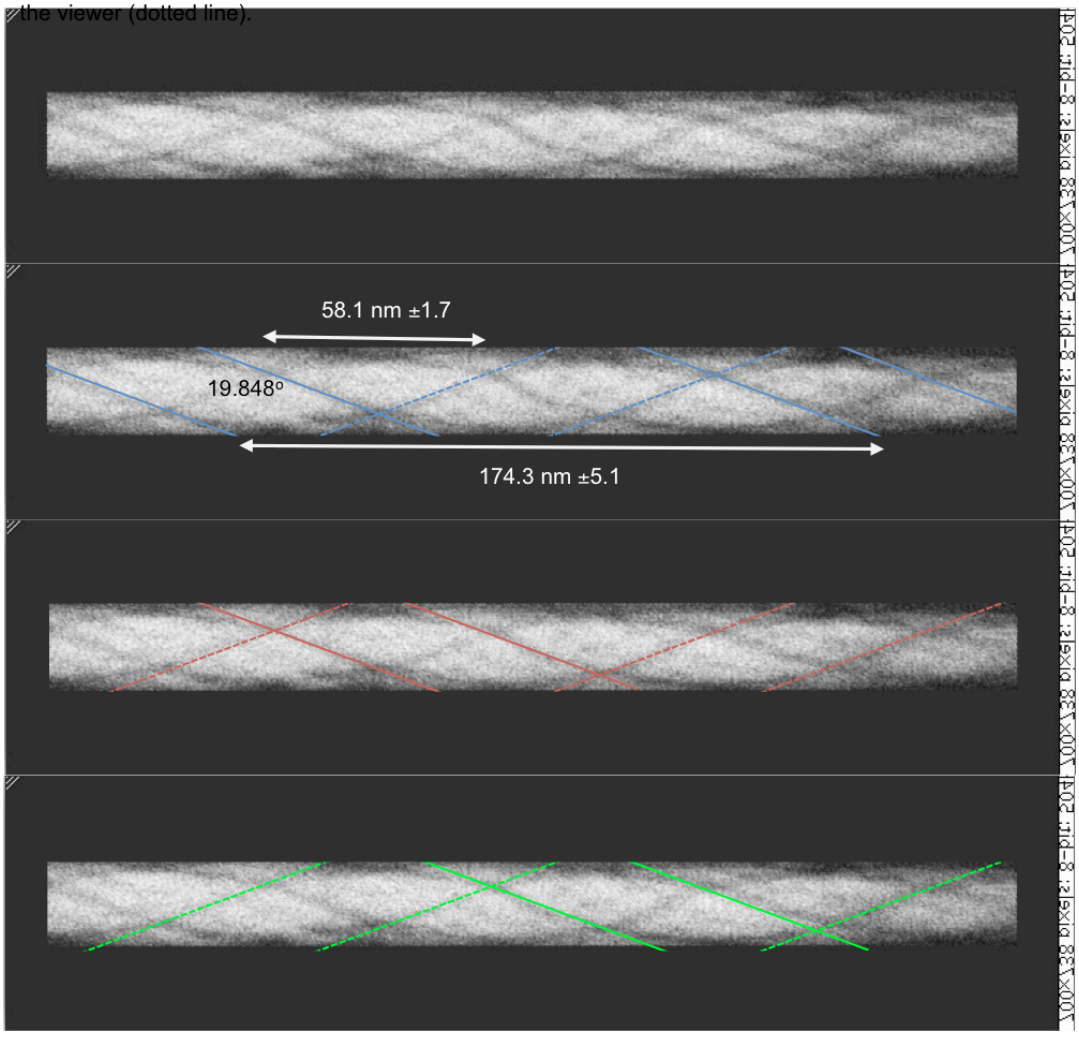
tan θ = opp/adj

tan 33.45 = 21.03/adj

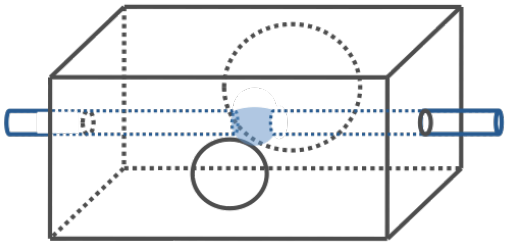
adj = 21.03/tan 33.45

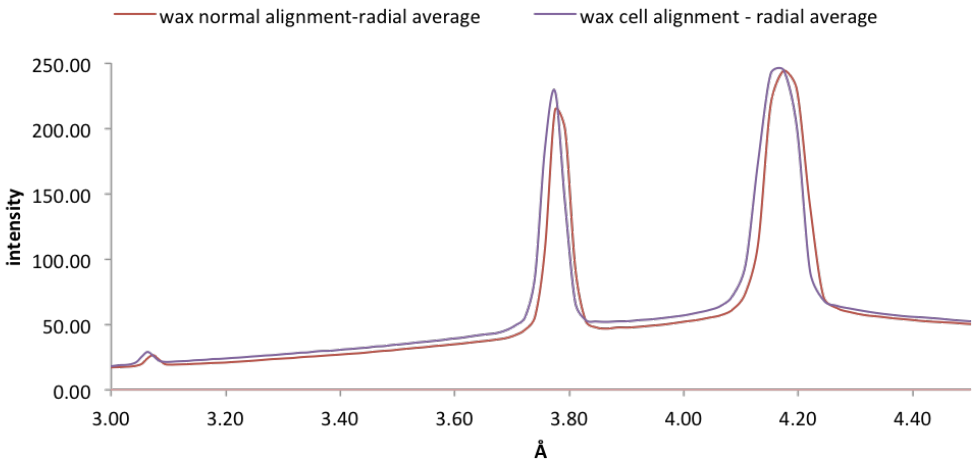
adj = 31.79 nm

Mature fibril appears to be three filaments (blue, red and green) twisted around each other with the same helicity (we cannot ascribe left or right handedness), interpretation is complicated by electron translucency and seeing the filaments when they are nearest the viewer (solid line) but also farthest from the viewer (dotted line).



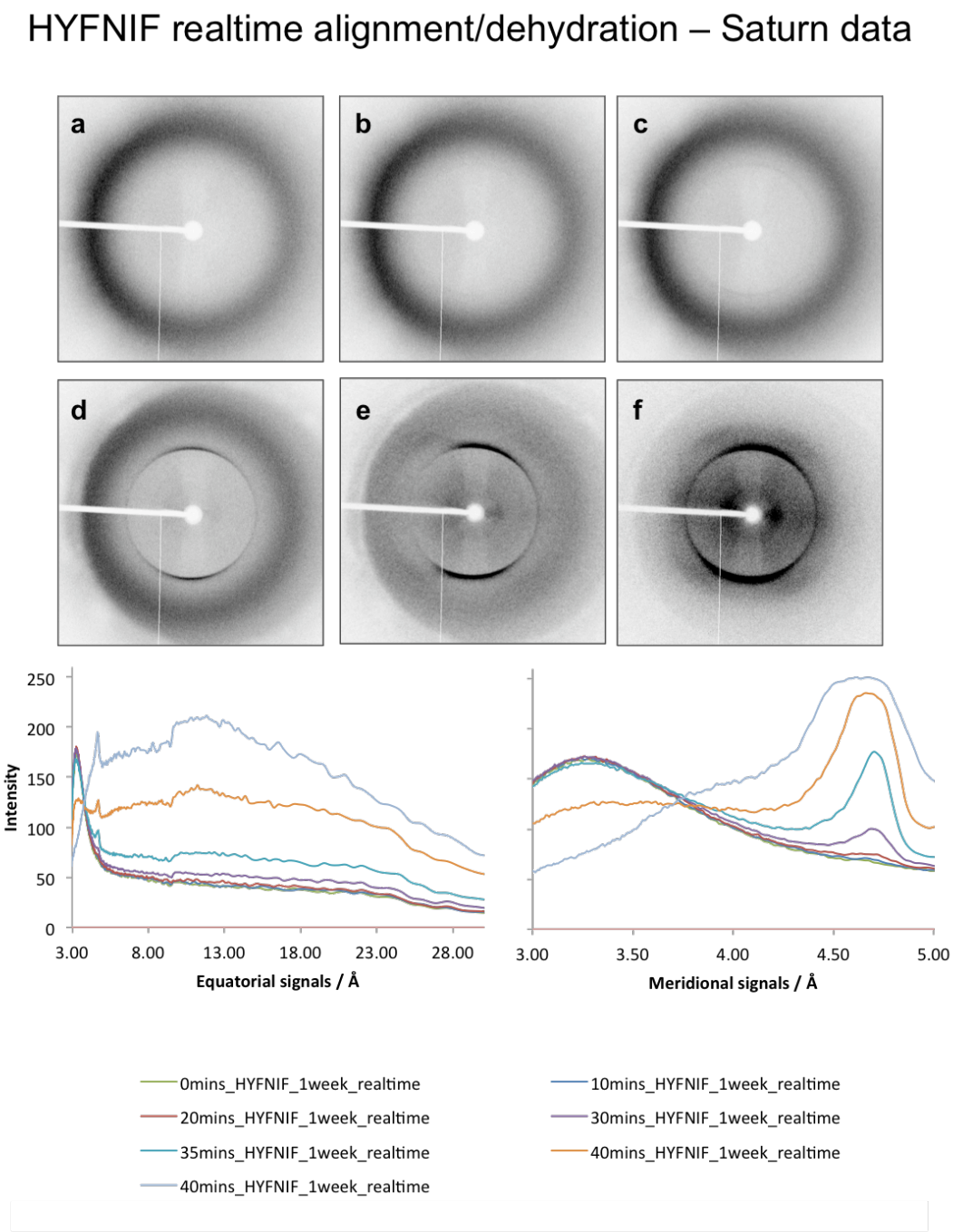
Realtime dehydration cell alignment/calibration check



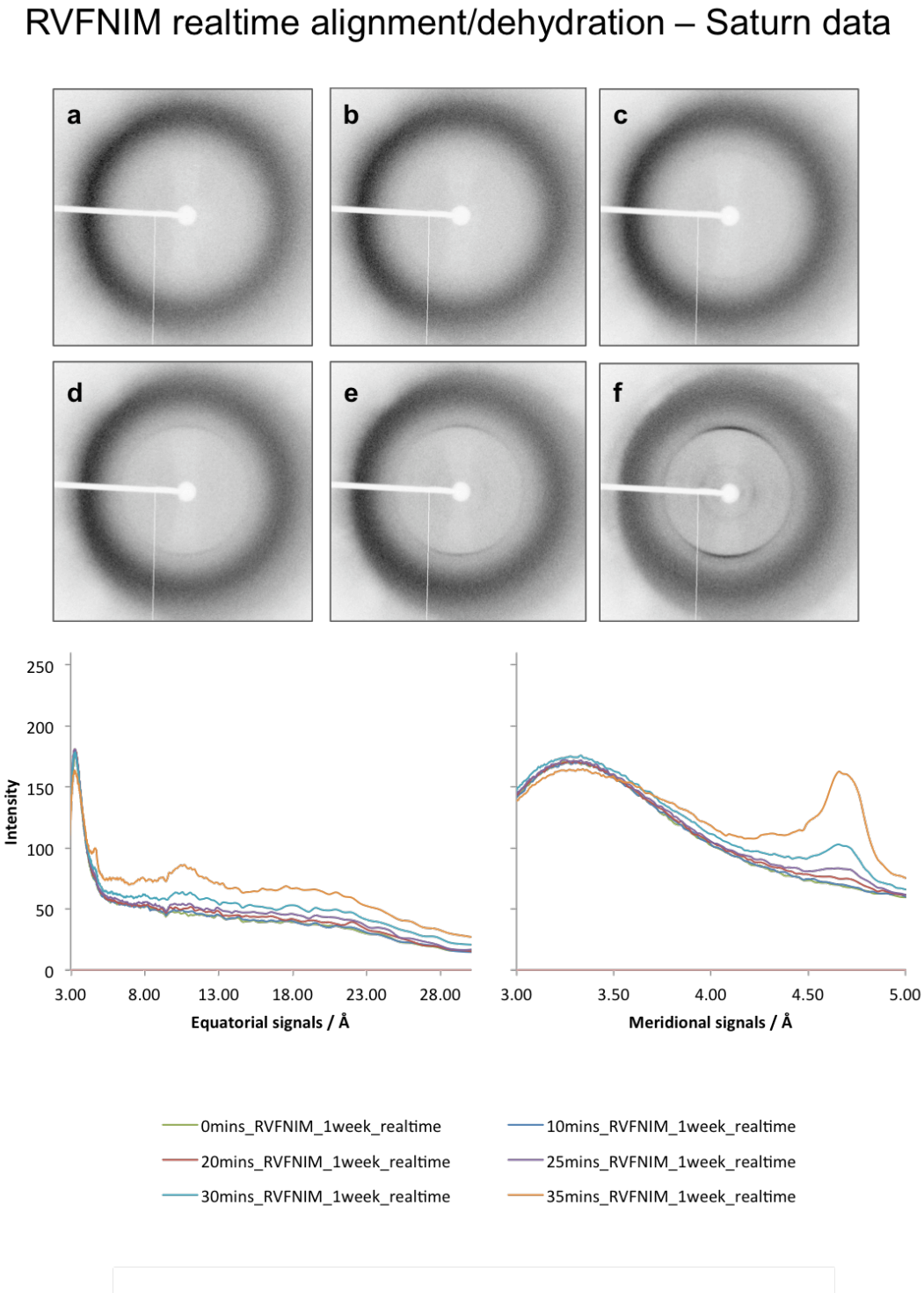


Graph showing the diffraction signals from a diffraction standard aligned manually by standard methods and using the dehydration cell. Peak intensities and positions are well matched. Despite the slight discrepancy observed graphically, measurement of the peaks finds them to be identical at 4.17 and 3.77 Å, thus any error in alignment leading to discrepancy in peak position in below the resolution of the patterns and is negligible.

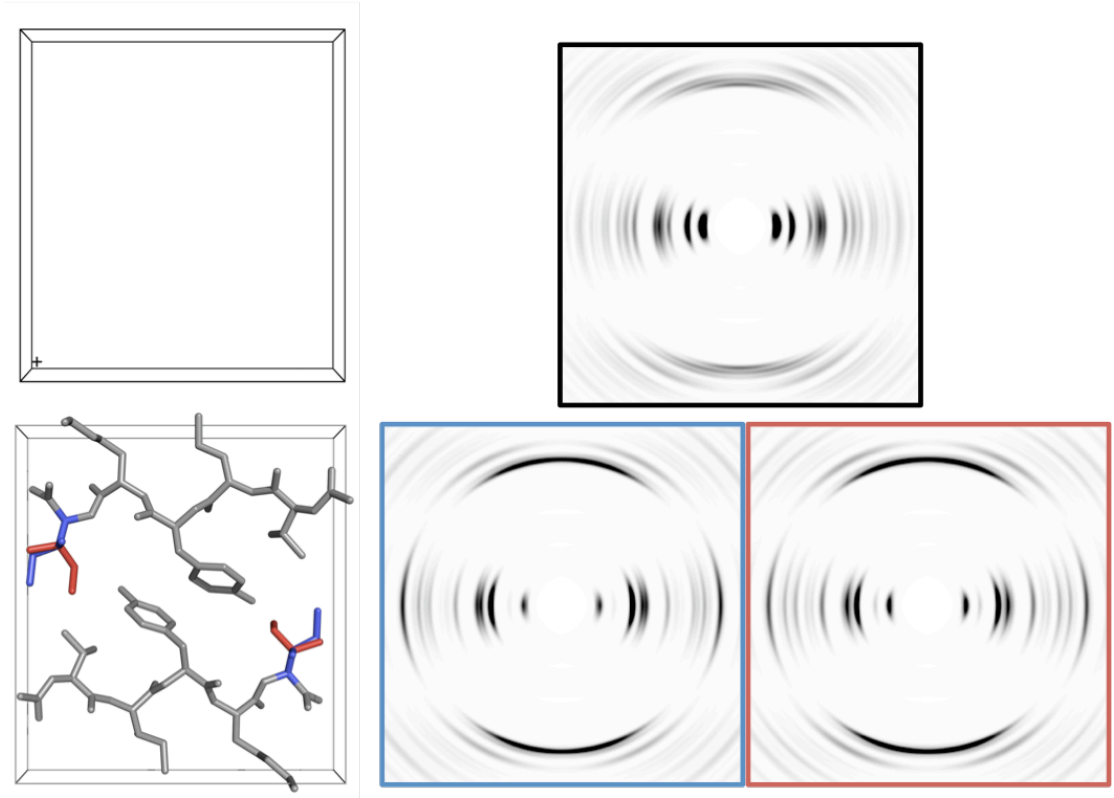
The major challenge with this piece of apparatus is the alignment of the sample, where the cameras standardly used to align samples cannot be used. Thus alignment must be inferred by the positioning of an adjacent pin aligned with the sample outside the dehydration cell.
It is vital to be able to align the sample accurately. Errors in sample alignment could lead to sample to detector distance errors and pattern artefacts. The above shows diffraction from a control sample with and without the dehydration chamber demonstrating the ability to successfully align samples by this method.



a-0mins, b-10mins, c-20mins, d-35mins-e-40mins,f-45mins

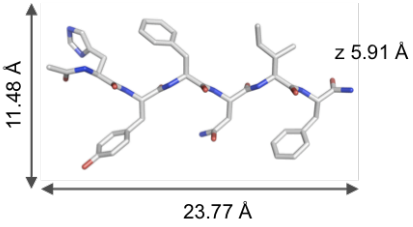

RVFNIM realtime alignment/dehydration – Saturn data
a
b
c
d
e
f
Intensity
250
200
150
100
50
0
3.00
8.00
13.00
18.00
23.00
28.00
Equatorial signals / Å
3.00
3.50
4.00
4.50
5.00
Meridional signals / Å
0mins_RVFNIM_1week_realtime
10mins_RVFNIM_1week_realtime
20mins_RVFNIM_1week_realtime
25mins_RVFNIM_1week_realtime
30mins_RVFNIM_1week_realtime
35mins_RVFNIM_1week_realtime

X-ray fibre diffraction simulation - signal modulation

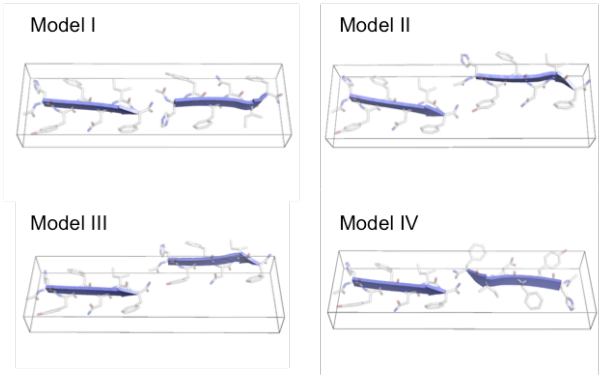


Waltz peptide monomer dimensions and minimised cell constructs

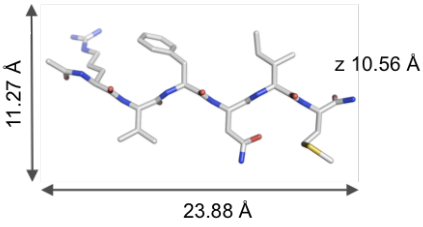
CH_3CONH – HYFNIF - $CONH_2$



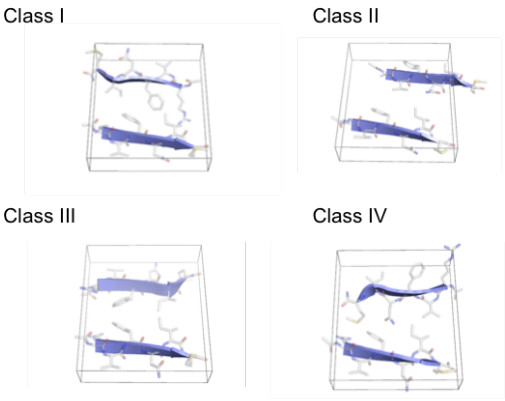
Van der Waals volume:
Monomer: 843.3 Å^3
Dimer: ~1712 Å^3
Unit cell: 2470 Å^3



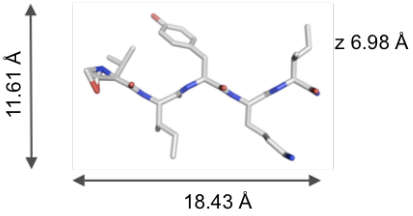
CH_3CONH – RVFNIM - $CONH_2$



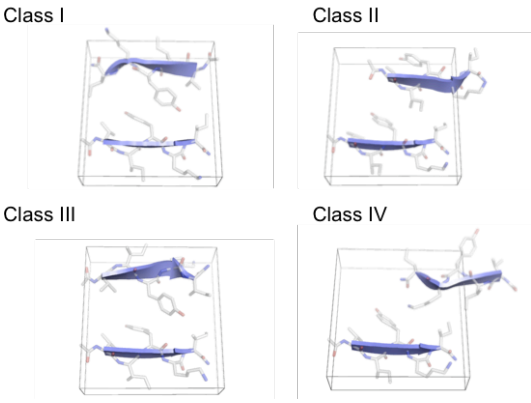
Van der Waals volume:
Monomer: 779.5 Å^3
Dimer: ~1580 Å^3
Unit cell: 1926 Å^3



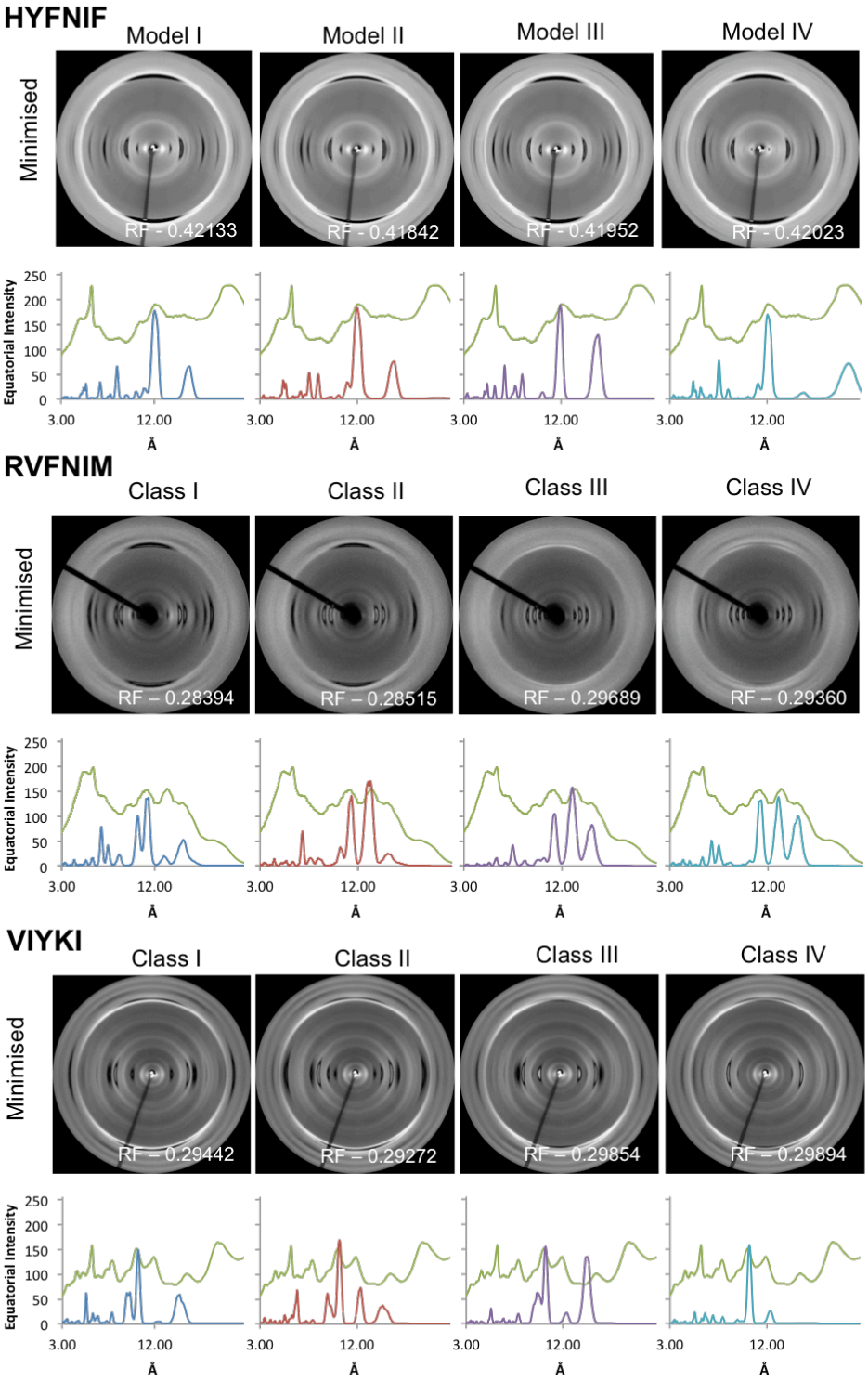
CH_3CONH – VIYKI - $CONH_2$



Van der Waals volume:
Monomer: 638.3 Å^3
Dimer: ~1354 Å^3
Unit cell: 1548 Å^3

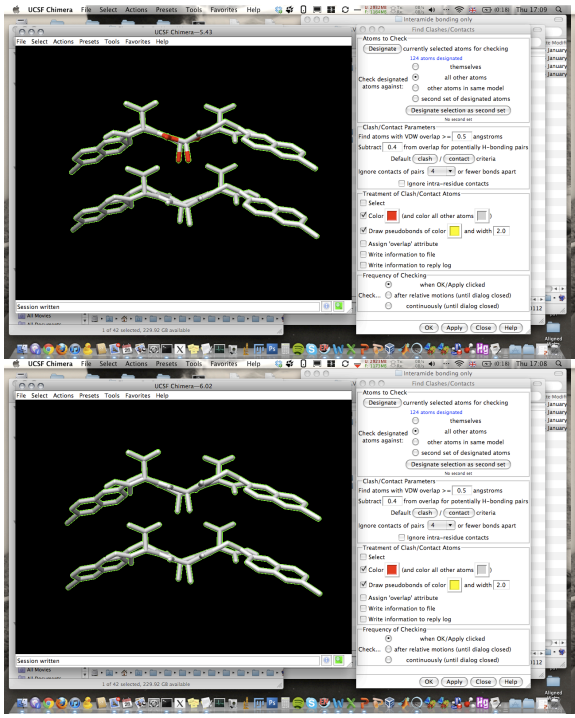


Waltz qualitative and quantitative comparison of simulated and experimental XRFD

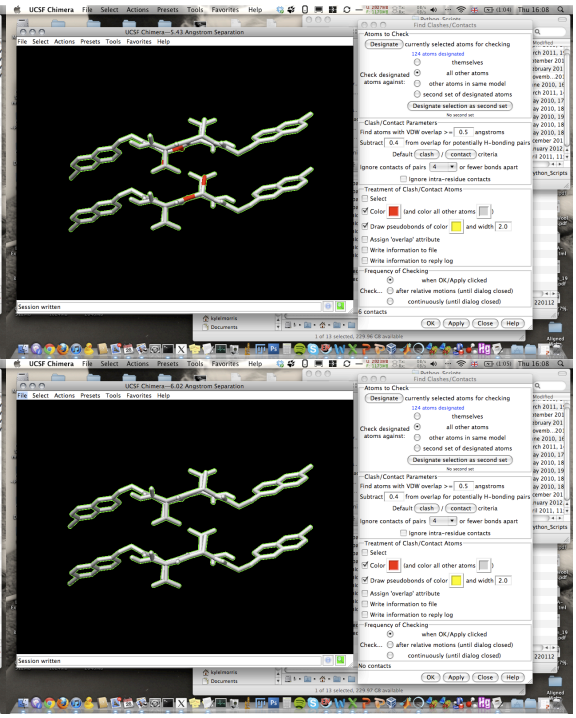


Appendix II

BrAV side chain packing – clash checks

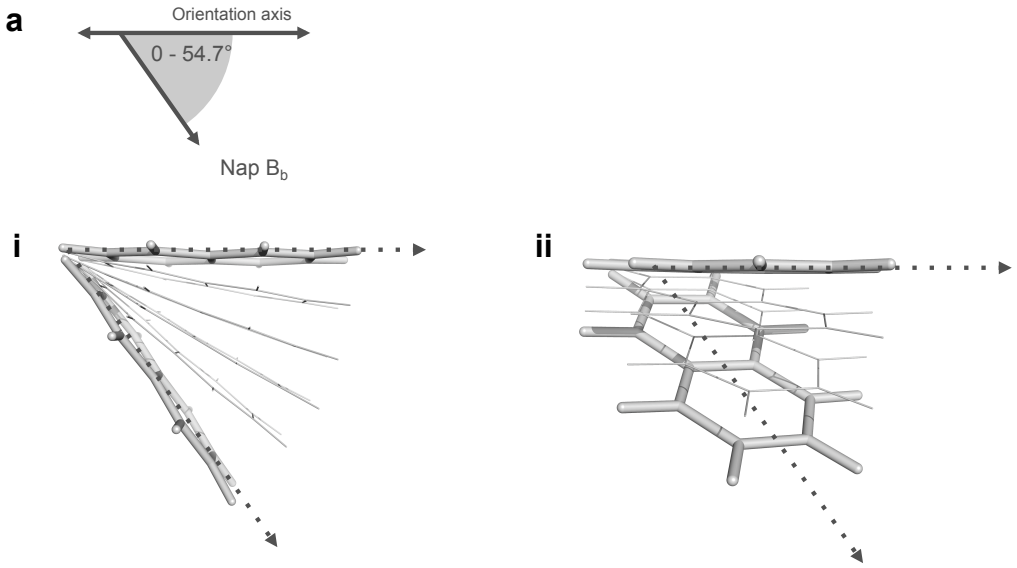


Antiparallel interamide
5.43 Å (top) and 6.03 Å (bottom) separation



Antiparallel interamide & interamide/carboxyl
5.43 Å (top) and 6.03 Å (bottom) separation

BrAV naphthalene orientation – LD measurement indications



Determined by LD measurements the long axes of the naphthalene groups lie less than 54.7° relative to the fibre axis {Chen 2010} (ai), also however, rotation of the long axis of the naphthalene group above and below the plane of the paper affects the angle through which the group might be found (aii).

BrAV unit cell indexing

Predicted cell	a37.8, b4.51, c27.2, α=β=γ=90
Search limit	a40.0, b32.0, c4.51, α=β=γ=90
% Error	0.0243

Observed	[h k l]	Predicted
38.00	[1 0 0]	37.79
27.00	[0 0 1]	27.23
21.00	[1 0 1]	22.09
16.00	[2 0 1]	15.52
13.00	[1 0 2]	12.81
11.00	[2 0 2]	11.05
7.30	[5 0 1]	7.28
6.50	[4 0 3]	6.54
6.02	[3 0 4]	5.99
5.40	[7 0 0]	5.40
5.00	[3 0 5]	5.00

Predicted cell	a54.3, b4.51, c37.0, α=β=γ=90
Search limit	a66.0, b40.0, c4.51, α=β=γ=90
% Error	0.0420

Observed	[h k l]	Predicted
65.00	[1 0 0]	54.28
38.00	[0 0 1]	37.03
27.00	[2 0 0]	27.14
21.00	[2 0 1]	21.89
16.00	[3 0 1]	16.26
13.00	[3 0 2]	12.94
11.00	[4 0 2]	10.95
7.30	[6 0 3]	7.30
6.50	[4 0 5]	6.50
6.02	[2 0 6]	6.02
5.43	[10 0 0]	5.43
5.00	[8 0 5]	5.00

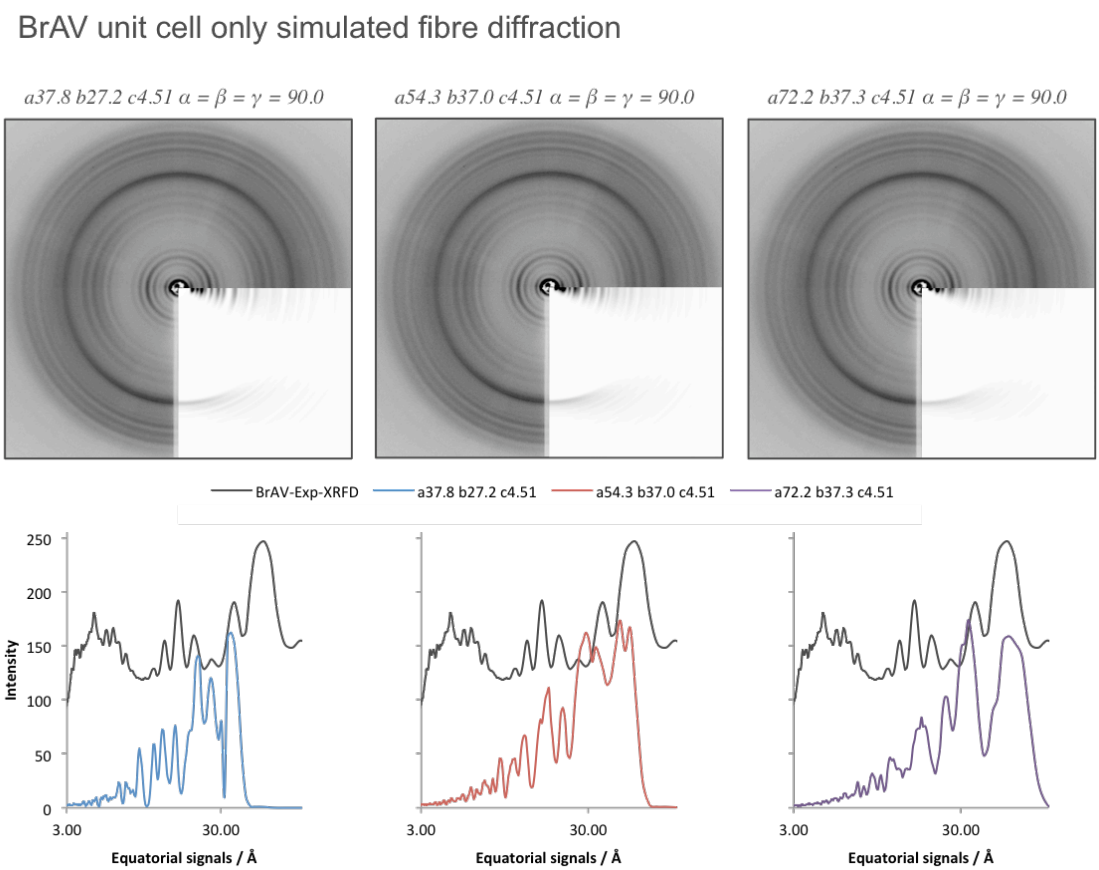
Predicted cell	a72.2, b4.51, c37.3, α=β=γ=90
Search limit	a80.0, b40.0, c4.51, α=β=γ=90
% Error	0.0262

Observed	[h k l]	Predicted
65.00	[1 0 0]	72.45
38.00	[0 0 1]	37.29
27.00	[2 0 1]	25.98
21.00	[3 0 1]	20.27
16.00	[4 0 1]	16.29
13.00	[4 0 2]	12.99
11.00	[3 0 3]	11.05
7.30	[2 0 5]	7.30
6.50	[8 0 4]	6.50
6.02	[3 0 6]	6.02
5.43	[12 0 3]	5.43
5.00	[5 0 7]	5.00

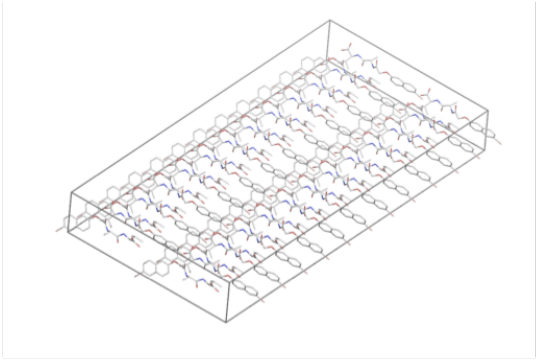
				Total standard error			Total upper error			Total lower error		Calculated value within experimental error
a37.8, b4.51, c27.2, α90, ß90, γ90				0.018			0.033			0.026		
h	**k**	**l**	**Calc / Å**	**Observed / Å**	**Error**	**+d / Å**	**Upper**	**Error**	**-d / Å**	**Lower**	**Error**	**Yes / No**
1	0	0	37.80	38.34	0.01	2.39	40.73	0.07	2.12	36.22	0.04	Yes
0	0	1	27.20	27.49	0.01	1.98	29.47	0.08	1.73	25.76	0.06	Yes
1	0	1	22.08	20.70	0.07	0.61	21.31	0.04	0.57	20.13	0.10	No
2	0	1	15.52	16.11	0.04	0.33	16.44	0.06	0.32	15.79	0.02	No
1	0	2	12.80	13.12	0.02	0.18	13.30	0.04	0.18	12.94	0.01	No
2	0	2	11.04	11.10	0.01	0.16	11.26	0.02	0.15	10.95	0.01	Yes
5	0	1	7.28	7.28	0.00	0.05	7.33	0.01	0.05	7.23	0.01	Yes
4	0	3	6.54	6.49	0.01	0.06	6.55	0.00	0.06	6.43	0.02	Yes
3	0	4	5.98	6.02	0.01	0.04	6.06	0.01	0.04	5.98	0.00	Yes
7	0	0	5.40	5.43	0.01	0.03	5.46	0.01	0.03	5.40	0.00	Yes

				Total standard error			Total upper error			Total lower error		Calculated value within experimental error
a54.3, b4.51, c37, α90, ß90, γ90				0.029			0.055			0.022		
h	**k**	**l**	**Calc / Å**	**Measured / Å**	**Error**	**+d / Å**	**Upper**	**Error**	**-d / Å**	**Lower**	**Error**	**Yes / No**
1	0	0	54.30	65.27	0.17	14.27	79.54	0.32	9.93	55.34	0.02	No
0	0	1	37.00	38.34	0.03	2.39	40.73	0.09	2.12	36.22	0.02	Yes
2	0	0	27.15	27.49	0.01	1.98	29.47	0.08	1.73	25.76	0.05	Yes
2	0	1	21.89	20.70	0.06	0.61	21.31	0.03	0.57	20.13	0.09	No
3	0	1	16.26	16.11	0.01	0.33	16.44	0.01	0.32	15.79	0.03	Yes
3	0	2	12.94	13.12	0.01	0.18	13.30	0.03	0.18	12.94	0.00	No
4	0	2	10.94	11.10	0.01	0.16	11.26	0.03	0.15	10.95	0.00	No
6	0	3	7.30	7.28	0.00	0.05	7.33	0.00	0.05	7.23	0.01	Yes
4	0	5	6.50	6.49	0.00	0.06	6.55	0.01	0.06	6.43	0.01	Yes
2	0	6	6.01	6.02	0.00	0.04	6.06	0.01	0.04	5.98	0.01	Yes
10	0	0	5.43	5.43	0.00	0.03	5.46	0.01	0.03	5.40	0.01	Yes

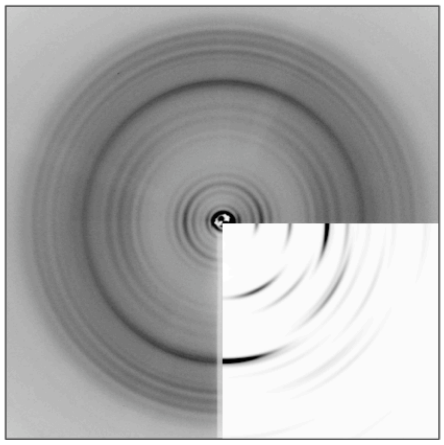
				Total standard error			Total upper error			Total lower error		Calculated value within experimental error
a72.2, b4.51, c37.3, α90, ß90, γ90				0.022			0.039			0.038		
h	**k**	**l**	**Calc / Å**	**Measured / Å**	**Error**	**+d / Å**	**Upper**	**Error**	**-d / Å**	**Lower**	**Error**	**Yes / No**
1	0	0	72.20	65.27	0.11	14.27	79.54	0.09	9.93	55.34	0.30	Yes
0	0	1	37.30	38.34	0.03	2.39	40.73	0.08	2.12	36.22	0.03	Yes
2	0	1	25.94	27.49	0.06	1.98	29.47	0.12	1.73	25.76	0.01	Yes
3	0	1	20.22	20.70	0.02	0.61	21.31	0.05	0.57	20.13	0.00	Yes
4	0	1	16.25	16.11	0.01	0.33	16.44	0.01	0.32	15.79	0.03	Yes
4	0	2	12.97	13.12	0.01	0.18	13.30	0.02	0.18	12.94	0.00	Yes
3	0	3	11.05	11.10	0.00	0.16	11.26	0.02	0.15	10.95	0.01	Yes
2	0	5	7.31	7.28	0.00	0.05	7.33	0.00	0.05	7.23	0.01	Yes
8	0	4	6.49	6.49	0.00	0.06	6.55	0.01	0.06	6.43	0.01	Yes
3	0	6	6.02	6.02	0.00	0.04	6.06	0.01	0.04	5.98	0.01	Yes
12	0	3	5.42	5.43	0.00	0.03	5.46	0.01	0.03	5.40	0.00	Yes


BrAV unit cell only simulated fibre diffraction
a37.8 b27.2 c4.51 α = β = γ = 90.0
a54.3 b37.0 c4.51 α = β = γ = 90.0
a72.2 b37.3 c4.51 α = β = γ = 90.0
BrAV-Exp-XRFD
a37.8 b27.2 c4.51
a54.3 b37.0 c4.51
a72.2 b37.3 c4.51
Intensity
250
200
150
100
50
0
3.00
30.00
Equatorial signals / Å

Initial modelling attempts

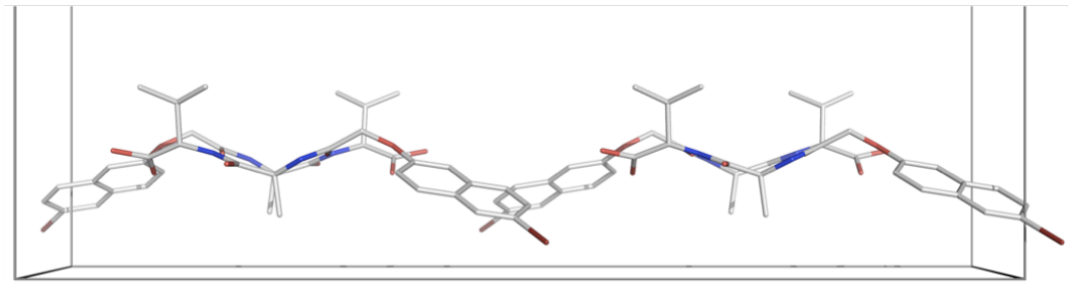


a 37.2 b 9.02 c 72.2 α 90.0 β 90.0 γ 90.0

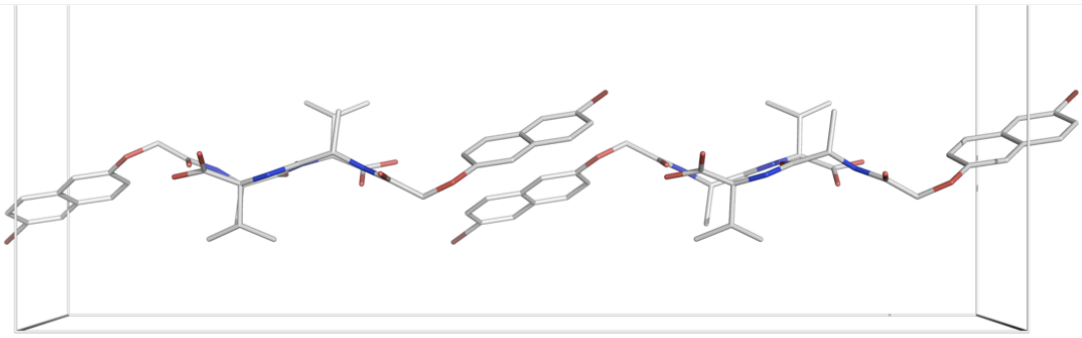


Unit cell modelling

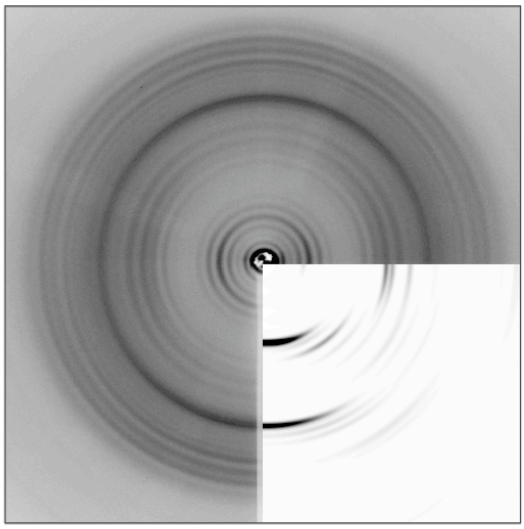
Interamide only hydrogen bonding

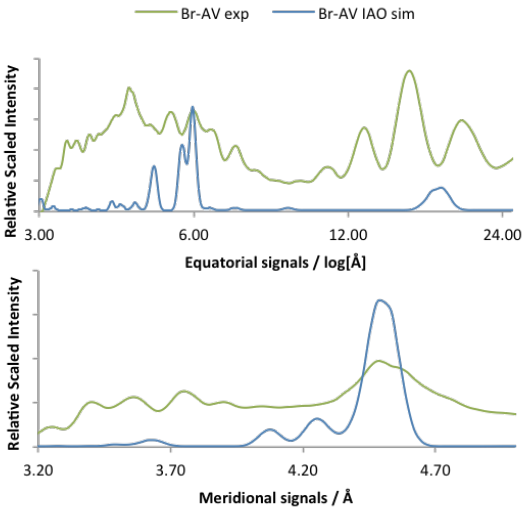


Interamide-carboxyl hydrogen bonding

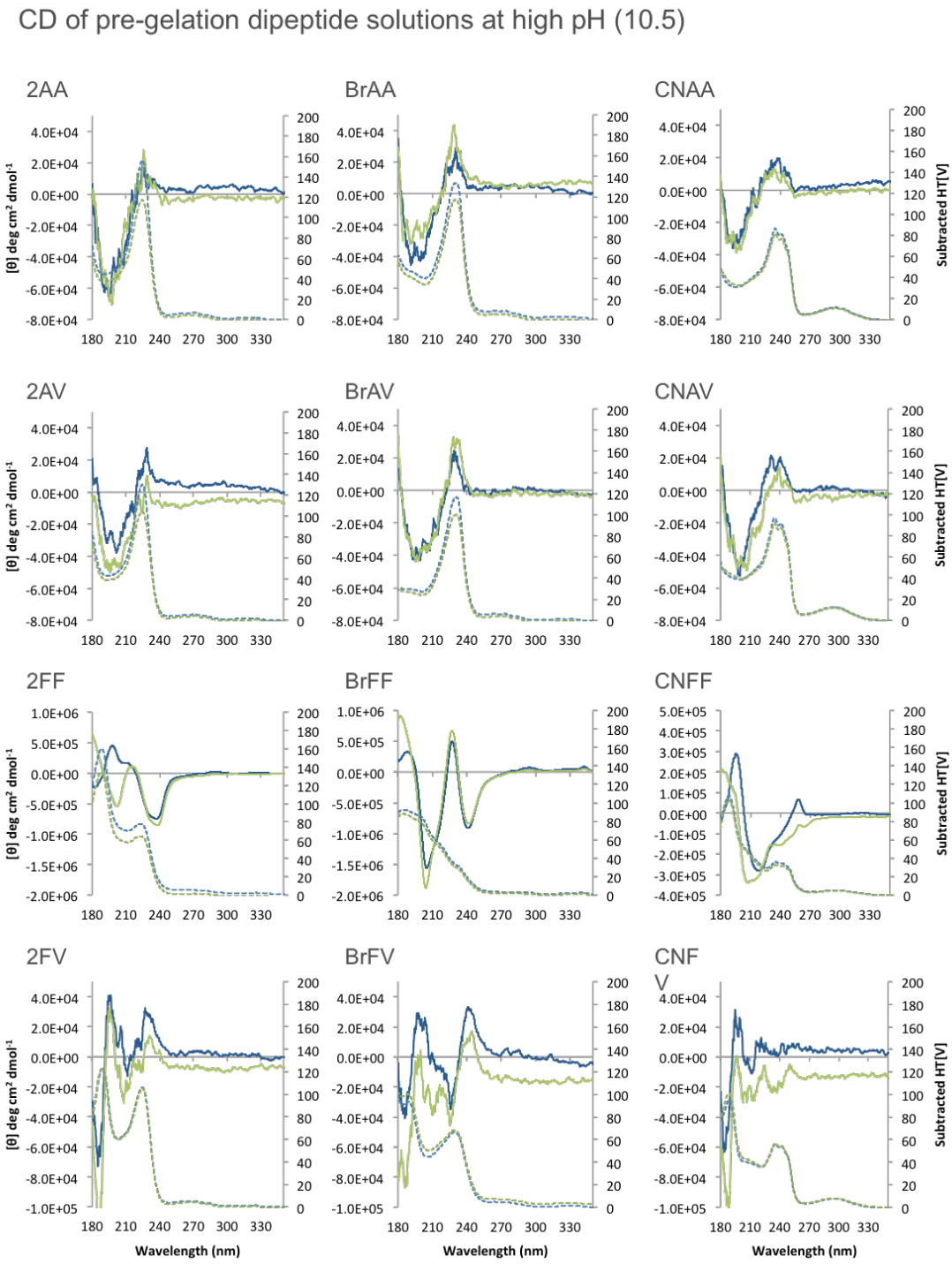


BrAV final model simulations





Appendix III



$HT[V]_{sub}$ of pre-gelation dipeptide solutions at high pH (10.5)

2AA		BrAA		CNAA	
$HT[V]_{sub}$ / nm	V_{sub}	$HT[V]_{sub}$ / nm	V_{sub}	$HT[V]_{sub}$ / nm	V_{sub}
224.5	155.78	230.4	133.46	234.8	86.48
262.4	6.22	263.3	8.57	241.2	81.04
270.9	6.57	273.0	8.40	295.1	11.45
~282	*	318.7	2.54		
310.5	1.66	332.8	2.70		
324.8	2.07				

2AV		BrAV		CNAV	
$HT[V]_{sub}$ / nm	V_{sub}	$HT[V]_{sub}$ / nm	V_{sub}	$HT[V]_{sub}$ / nm	V_{sub}
224.9	130.09	230.9	116.34	234.8	96.96
262.4	5.48	263.5	6.14	241.3	91.45
270.9	5.84	273.2	6.18	295.1	12.65
281.7	4.15	319.3	1.19		
310.5	1.72	332.8	1.31		
324.8	2.10				

2FF		BrFF		CNFF	
$HT[V]_{sub}$ / nm	V_{sub}	$HT[V]_{sub}$ / nm	V_{sub}	$HT[V]_{sub}$ / nm	V_{sub}
188.3	160.20	184.6	92.21	188.3	107.04
222.8	77.84	278.9	3.19	235.4	35.69
260.9	5.69	290.9	2.82	241.9	34.57
270.5	5.49	325.9	1.89	295.9	5.17
324.8	2.62	340.7	2.75		

2FV		BrFV		CNFV	
$HT[V]_{sub}$ / nm	V_{sub}	$HT[V]_{sub}$ / nm	V_{sub}	$HT[V]_{sub}$ / nm	V_{sub}
189.1	122.79	188.6	93.93	188.5	94.71
224.7	106.28	228.9	66.52	235.4	55.97
262.4	5.00	264.5	4.86	241.8	53.93
271.3	5.16	274.2	4.52	295.1	7.57
310.4	1.40	307.1	0.61		
324.7	1.59	321.9	1.08		
		335.4	1.30		

The material phases formed by the naphthalene dipeptide library

Photographs of excised gels visualised by cross-polarised light microscopy. The scale bars represent 2 mm.

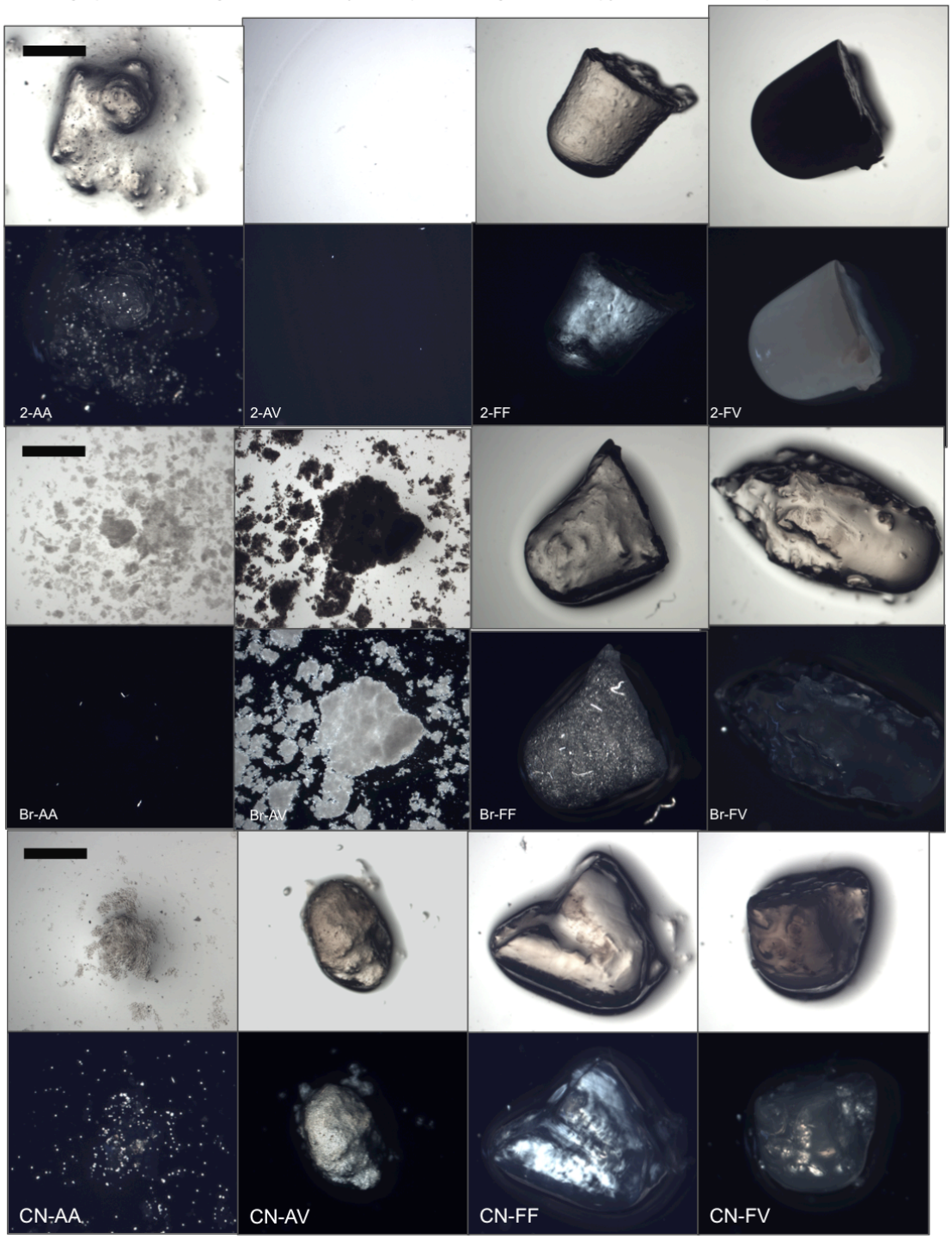


Table of CD and HT[V]$_{sub}$ signals

2-AA 110311

CD

0 mins		120 mins		End (350 mins)	
Peaks / nm	[θ]	Peaks / nm	[θ]	Peaks / nm	[θ]
213.3	-7.92E+04	196.5	3.76E+05	195.9	5.69E+05
		207.8	-6.77E+05	207.7	-9.42E+05
		220.9	4.98E+05	221.1	7.94E+05
		233.2	4.18E+05	*235	6.00E+05
		279.3	4.93E+04	280.3	6.93E+04
		291.0	6.67E+04	291.4	9.36E+04

Sub HT

0 mins		120 mins		End (350 mins)	
Peaks / nm	HT_{sub}/V	Peaks / nm	HT_{sub}/V	Peaks / nm	HT_{sub}/V
				181.7	106.2
				199.5	84.0
224.2	163.7	223.1	94.2	*220	*71.7
262.3	8.9	261.6	9.0		
270.8	9.4	270.1	9.1		
*283	*6.5			277.8	8.8
		*292	*6.6	290.5	8.5
311.5	3.0	310.2	4.2		
		314.3	4.4	315.1	5.0

2AV 150811

CD

0 mins		120 mins		End (250 mins)	
Peaks / nm	[θ]	Peaks / nm	[θ]	Peaks / nm	[θ]
211.0	-7.91E+04	208.9	-6.20E+04	212.8	-5.26E+04

Sub HT

0 mins		120 mins		End (250 mins)	
Peaks / nm	HT_{sub}/V	Peaks / nm	HT_{sub}/V	Peaks / nm	HT_{sub}/V
		182.3	70.8	182.7	58.2
224.5	174.7	224.5	97.6	224.5	63.5
		251.2*	10.2	250.8*	9.1
263.0	6.5	260.7	10.6	262.3	9.4
270.4	7.0	270.2	10.7	270.4	9.6
283*	4.5	283*	9.3	283*	8.8
310.1	1.9	308.9	7.2	308.9	7.8
324.1	2.4	324.0	7.1	324.0	7.7

2-FF (ThT) 220711

CD

0 mins		120 mins		End (270 mins)	
Peaks / nm	[θ]	Peaks / nm	[θ]	Peaks / nm	[θ]
194.5	3.70E+05	193.7	2.19E+05	193.0	1.79E+05
205.5	-3.16E+05	207.0	-3.29E+05	207.0	-2.52E+05
212.7 (+)	-1.21E+05	224.4	4.96E+05	225.1	4.50E+05
236.2	-1.14E+06	239.6	-2.89E+05	239.9	-2.13E+05
265.3	-8.66E+04				
275.5	-4.81E+04				
290.5	2.30E+03	289.2	7.02E+03	291.3	7.61E+03

Sub HT

0 mins		120 mins		End (270 mins)	
Peaks / nm	HT_{sub}/V	Peaks / nm	HT_{sub}/V	Peaks / nm	HT_{sub}/V
<180	209.0	184.7	88.7	188.4	57.2
276.8	4.6	276.8	5.6	276.8	7.6
288.4	4.2	288.5	5.4	288.3	7.4
*310	2.4				
319.9	2.7	317.6	3.6	316.0	5.5
325.7	2.7				
332.2	3.0	331.4	3.9	331.4	5.6

2-FV (ThT) 080711

CD

0 mins		120 mins		End (200 mins)	
Peaks / nm	[θ]	Peaks / nm	[θ]	Peaks / nm	[θ]
195.2	9.03E+04	193.2	2.74E+05	191.9	2.80E+05
				203.3 (-)	9.84E+04
		216.3	2.61E+05	216.7	2.68E+05
234.1	-1.25E+05	235.1	-7.18E+05	235.0	-6.87E+05
		288.0	3.49E+04	288.1	3.73E+04
				316.2	1.20E+04
				330.4	1.39E+04

Sub HT

0 mins		120 mins		End (200 mins)	
Peaks / nm	HT_{sub}/V	Peaks / nm	HT_{sub}/V	Peaks / nm	HT_{sub}/V
187.9	124.8	184.0	76.5	184.0	74.0
224.5	96.2	220.0	53.5	219.4	51.9
262.7	7.9			*265	8.7
270.4	8.0	274.6	8.1	274.5	8.4
*285	6.0			*285	6.9
309.9	3.9	315.4	4.2	315.6	4.3
325.1	4.0	325.8	4.0	325.9	4.1

Br-AA 180311

CD

0 mins		120 mins		End (270 mins)	
Peaks / nm	[θ]	Peaks / nm	[θ]	Peaks / nm	[θ]
214.3	-8.06E+04	202.5	5.15E+05	202.6	5.06E+05
		212.0	-5.67E+05	211.8	-5.84E+05
		245.0	1.87E+05	245.5	1.97E+05
		298.8	3.86E+04	297.7	4.40E+04

Sub HT

0 mins		120 mins		End (270 mins)	
Peaks / nm	HT_{sub}/V	Peaks / nm	HT_{sub}/V	Peaks / nm	HT_{sub}/V
		182.2	43.6	182.7	39.9
		200.7	26.7	200.7	25.8
230.4	106.2	228.9	19.9	226.9	15.3
263.4	5.8				
272.6	5.8				
*284	*3.5	282.3	3.1	282.4	3.3
		297.9	3.0	296.9	3.5

Br-AV 040311

CD

0 mins		120 mins		End (240 mins)	
Peaks / nm	[θ]	Peaks / nm	[θ]	Peaks / nm	[θ]
216.8	-7.97E+04	192.0	6.99E+05	192.0	7.36E+05
		211.2	-1.15E+06	211.4	-1.17E+06
		235.5	-2.09E+06	235.7	-2.23E+06
		276.5	-2.58E+05	276.5	-2.79E+05
		287.0	-2.09E+05	287.2	-2.21E+05

Sub HT

0 mins		120 mins		End (240 mins)	
Peaks / nm	HT_{sub}/V	Peaks / nm	HT_{sub}/V	Peaks / nm	HT_{sub}/V
230.7	97.1	224.0	49.1	223.4	48.7
263.3	5.7	268.0	5.3	*268	*5.5
272.5	5.6	278.0	4.5	*278	*4.7
*285	*3.16				

Br-FF (ThT) 150711

CD

0 mins		120 mins		End (230 mins)	
Peaks / nm	[θ]	Peaks / nm	[θ]	Peaks / nm	[θ]
189.2	2.28E+05	190.4	3.21E+05	189.8	3.38E+05
211.5	-7.80E+05	210.9	-8.20E+05	210.3	-8.54E+05
227.3	1.99E+05	227.4	1.54E+05	227.1	1.94E+05
240.5	-5.54E+05	239.9	-6.74E+05	239.7	-7.09E+05
295.8	4.38E+04	291.7	4.34E+04	291.1	4.27E+04
325.9	3.03E+04	326.5	3.09E+04	326.5	3.15E+04
341.0	5.15E+04	341.5	5.45E+04	341.5	5.90E+04

Sub HT

0 mins		120 mins		End (230 mins)	
Peaks / nm	HT_{sub}/V	Peaks / nm	HT_{sub}/V	Peaks / nm	HT_{sub}/V
<180	90.0	184.3	74.1	185.7	66.2
*200	55.3	200.0	54.5	*200	55.9
325.7	1.7	325.7	2.2	325.7	3.1
340.2	2.1	340.2	2.5	340.2	3.3

Br-FV 310112

CD

0 mins		120 mins		End (200 mins)	
Peaks / nm	[θ]	Peaks / nm	[θ]	Peaks / nm	[θ]
196.3	6.76E+04	198.1	6.53E+04	196.2	6.35E+04
228.3	-2.82E+05	229.8	-2.16E+05	230.1	-1.97E+05
240.6	9.36E+04	251.1	9.21E+04	252.0	8.18E+04
		266.9 (-)	5.29E+04	266.9 (-)	4.57E+04
		276.1	5.62E+04	276.2 (-)	4.69E+04
		293.0 (-)	1.27E+05	293.0	1.16E+05
		306.9	1.02E+05		
		318.1 (-)	1.01E+05	318.3 (-)	9.11E+04
		325.9	1.20E+05	326.4	1.09E+05
		335.5 (-)	1.13E+05	335.4 (-)	1.04E+05
		343.1	1.28E+05	343.7	1.21E+05

Sub HT

0 mins		120 mins		End (200 mins)	
Peaks / nm	HT_{sub}/V	Peaks / nm	HT_{sub}/V	Peaks / nm	HT_{sub}/V
182.8	36.7	183.2	32.1		
231.0	13.1	230.4	12.4		
267.0	5.2				
322.8	3.5				
338.4	3.2				

CN-AA 310112

0 mins		120 mins		End (300 mins)	
Peaks / nm	[θ]	Peaks / nm	[θ]	Peaks / nm	[θ]
221.9	-4.25E+04	220.4	-1.85E+04	220.4	-1.36E+04

sub HT

0 mins		120 mins		End (300 mins)	
Peaks / nm	HT_{sub}/V	Peaks / nm	HT_{sub}/V	Peaks / nm	HT_{sub}/V
234.8	84.4	235.1	22.4	235.2	15.3
241.2	78.9	241.6	21.6	241.2	14.8
248.8	59.1				
295.1	11.2	294.8	4.3	295.2	3.3

CN-AV 240311

CD

0 mins		120 mins		End (140 mins)	
Peaks / nm	[θ]	Peaks / nm	[θ]	Peaks / nm	[θ]
		190.7	1.78E+05	190.1	1.83E+05
212.7	-5.25E+04	210.9	-2.70E+05	211.0	-2.80E+05
		256.3	-4.88E+05	256.4	-5.11E+05

Sub HT

0 mins		120 mins		End (140 mins)	
Peaks / nm	HT_{sub}/V	Peaks / nm	HT_{sub}/V	Peaks / nm	HT_{sub}/V
<180		183.2	43.0	183.3	41.9
234.6	64.4	234.2	37.7	234.1	35.6
241.2	61.7				
293.3	10.4	289.6	8.3	277.4	7.7

CN-FF 220611

CD

0 mins		120 mins		End (150 mins)	
Peaks / nm	[θ]	Peaks / nm	[θ]	Peaks / nm	[θ]
196.2	5.08E+05	197.3	6.05E+05	197.4	6.27E+05
221.5	-4.32E+05	239.3	-7.57E+05	239.5	-7.63E+05
258.5	1.31E+05	*255	-5.30E+05		
293.2	-1.02E+04	292.3	-1.04E+05		
324.5	1.19E+04	336.8	-2.24E+04	337.8	-2.32E+04

Sub HT

0 mins		120 mins		End (150 mins)	
Peaks / nm	Sub HT / V	Peaks / nm	Sub HT / V	Peaks / nm	Sub HT / V
		185.4	42.8	185.5	42.2
*190	57.1	199.9	40.5	200.3	40.7
*206	35.6				
234.6	18.9	*234	16.4	*237	15.9
*245	17.1	*250	13.1	*255	11.7
299.7	2.3	299.8	2.2	299.7	2.2

CN-FV 310311

CD

0 mins		120 mins		End (230 mins)	
Peaks / nm	[θ]	Peaks / nm	[θ]	Peaks / nm	[θ]
196.7	4.43E+04	195.4	2.73E+05	196.0	2.96E+05
212.9	-8.07E+04	205.8	-4.06E+05	205.9	-4.07E+05
220.7	-7.20E+04	226.8	-8.18E+04	227 (+)	-7.88E+04
		234.6	-1.70E+05	235.1	-1.78E+05
250 (+)	-2.76E+03	246.8	6.09E+04	246.7	5.38E+04
264.8	-1.68E+04	260.2	-1.27E+05	261.7	-1.40E+05
306.2	7.05E+03	306.0	1.78E+05	305.5	1.92E+05

Sub HT

0 mins		120 mins		End (230 mins)	
Peaks / nm	HT_{sub}/V	Peaks / nm	HT_{sub}/V	Peaks / nm	HT_{sub}/V
185.8	60.9	184.7	48.0	184.7	47.0
		200.1	39.2	200.5	39.0
235.3	36.9	236.7	29.3	236.5	28.8
241.4	35.7				
295.3	7.6	292.5	7.5	292.4	7.6

Dipeptide FTIR data in the assembled state

Data collected and provided by Lin Chen (Liverpool University)

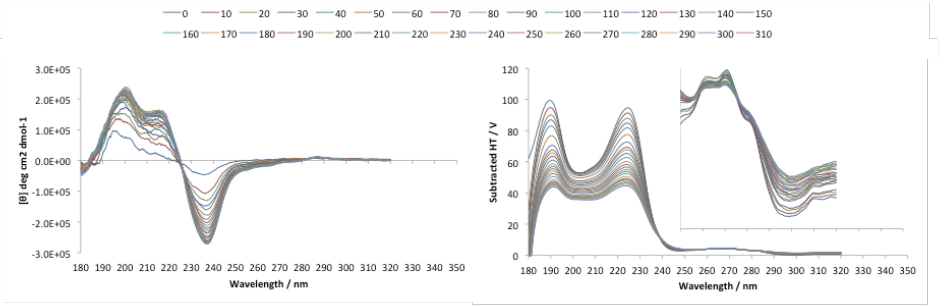
	Peaks / cm^{-1}						
	-COOH	Random coil / -C=C-	ß-sheet	Antiparallel ß-sheet			Relative % hydrogen bonding strength
System	1	2	3	4	Conformation	Phase	
2AA	1712.71	1644.24	1628.81	~1666.42	Antiparallel ß	Gel	79
2AV	1723.57		1629.37		Parallel ß	Gel	40
BrAA	1734.89	1651.71	1632.66		Parallel ß	Gel	0
BrAV	1720.86	1644.92	1627.48	~1680	Antiparallel ß	Gel	50
CNAA	1712.73		1626.68	1663.55	Antiparallel ß	Gel	79
CNAV	1733.76	1648.07	1625.91		Parallel ß	Gel	4
2FF	1716.61	1644.06	~1633	1663.94	Antiparallel ß	Gel	65
2FV	1706.92	1654.84	1628.81		Parallel ß	Gel	100
BrFF	1712.71	1643.27	1633.63	1664.45	Antiparallel ß	Gel	79
BrFV	1708.85	1651.95	1629.77	1659.67	Antiparallel ß	Crystal	93
CNFF	1712.71	~1641.34	~1634.59	~1664.49	Antiparallel ß	Gel	79
CNFV	1708.85	1654.84	1625.91	~1664.49	Antiparallel ß	Gel	93

FTIR dipeptide data (~indicates broad overlapping peaks)

Hydrogen bonding strength visualised as a gradient between blue (relatively weakest) to red (relatively strongest) by assuming the shortest wavenumbers for the –COOH group represent stronger hydrogen bonding patterns.

2-FG CD and XRFD data (colloidal system)

a



2-FG 070411
LD check No change

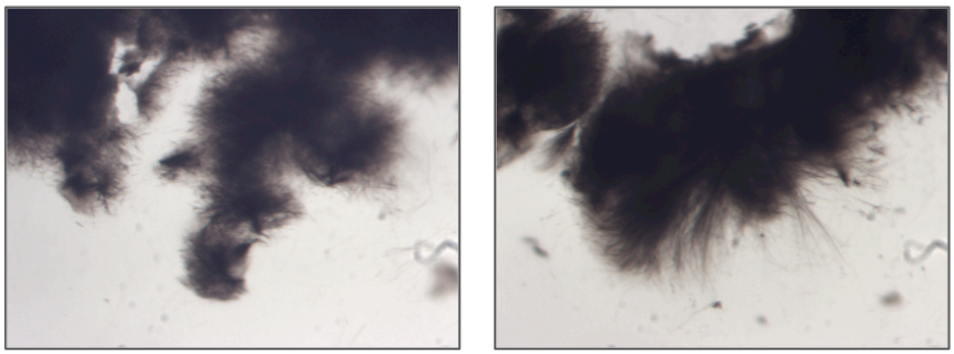
CD

0 mins		120 mins		End (240 mins)	
Peaks / nm	[θ]	Peaks / nm	[θ]	Peaks / nm	[θ]
194.7	9.73E+04	200.3	2.16E+05	200.6	2.30E+05
		210.6	1.48E+05	213.6	1.61E+05
235.4	-4.71E+04	236.9	-2.38E+05	236.9	-2.70E+05
		286.8	10501.76916	289.8	9.81E+03

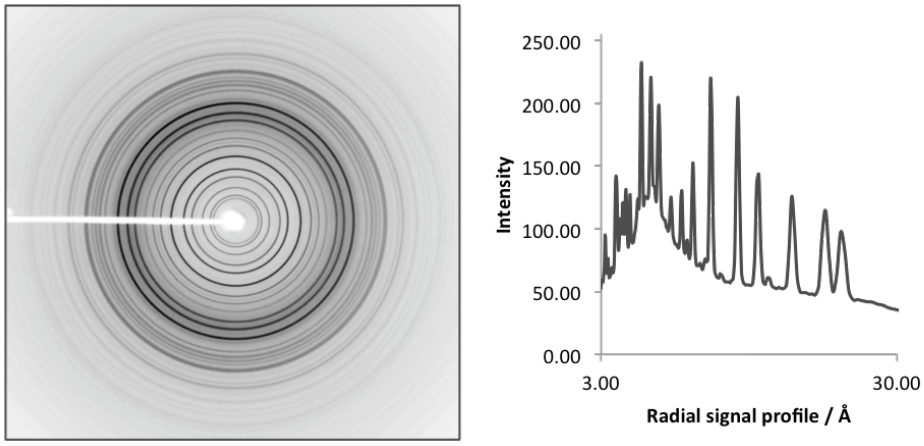
Sub HT

0 mins		120 mins		End (240 mins)	
Peaks / nm	Sub HT / V	Peaks / nm	Sub HT / V	Peaks / nm	Sub HT / V
189.5	99.341	191	57.062	191.4	43.895
224.5	94.618	224	58.262	223.7	44.634
263	4.171	263.4	4.025	261.6	3.984
270.7	4.485	270.5	4.149	271	4.054
*283	*2.82	*283	*2.91	*283	*3.07
311	0.847				

b

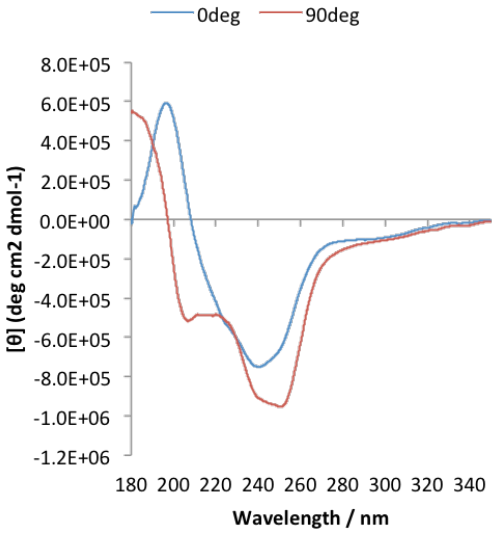


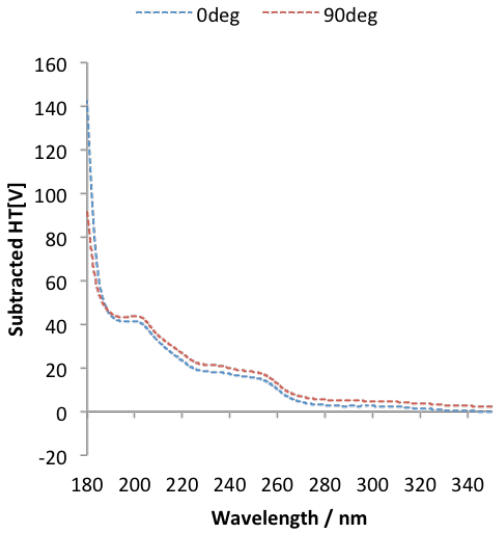
c



a) The circular dichroism exhibited by 2-FG upon self-assembly.
b) The system is seen to behave in a collodial manner, confirmed by the observation of spherulites
c) The spherulites have a high crystalline order and exhibit diffraction akin to powder diffraction.

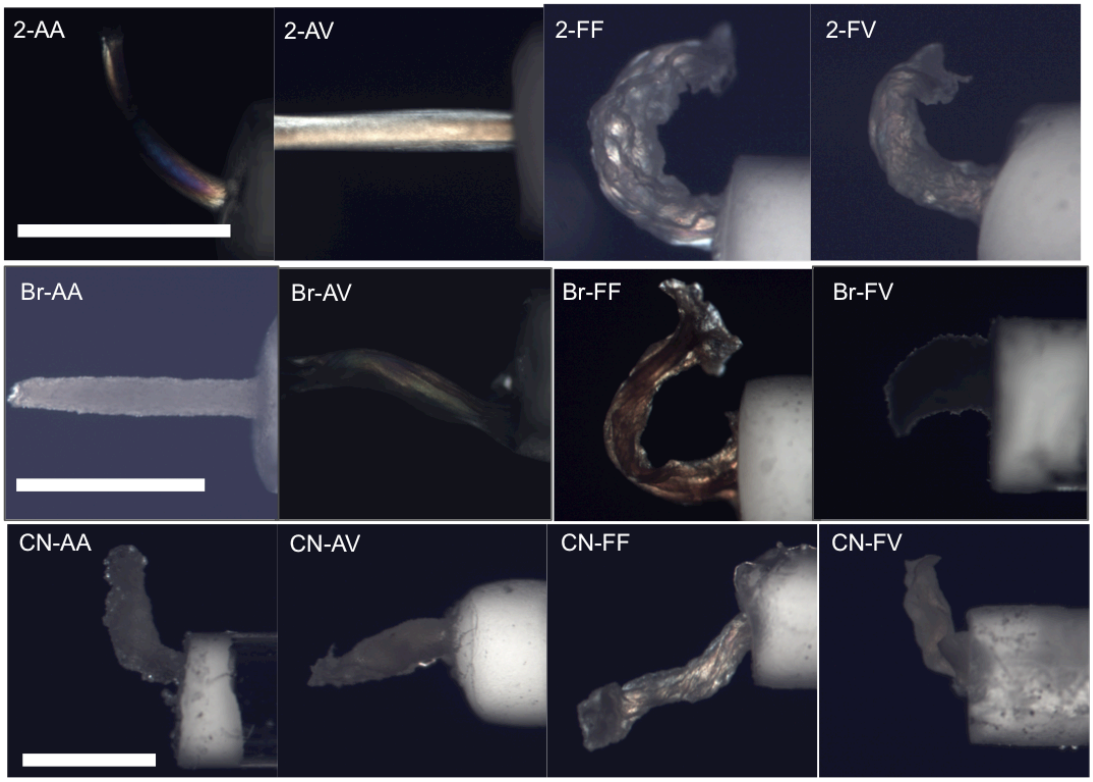
CN-FF CD cuvette rotation (LD artifacts)





Alignments of naphthalene dipeptide assemblies for X-ray fibre diffraction

Photographs of *in situ* alignments of visualised by cross-polarised light microscopy. The scale bars represent 1.2 mm.



Naphthalene dipeptide library fibre diffraction signal positions

2-AA

Axis	d-Spacings (Å)	Intensity	+d (Å)	-d (Å)
E	40.52	206.95	4.81	3.89
E	14.37	115.69	0.44	0.41
E	10.65	96.33	0.23	0.22
E	7.23	178.26	0.01	0.01
E	5.79	242.72	0.05	0.05
M	4.68	228.01	0.01	0.01
Rad (M)	4.27	177.89	0.00	0.00
Rad (M)	3.99	182.14	0.01	0.01
Rad (M)	3.38	138.50	0.01	0.01

Br-AA (Crystalline sampling)

Axis	d-Spacings (Å)	Intensity	+d (Å)	-d (Å)
E	33.38	249.70	0.57	0.59
E	16.86	120.12	0.34	0.35
E	11.18	86.49	0.16	0.17
E	8.38	75.04	0.08	0.08
E	6.61	73.11	0.06	0.07
M	5.43	245.33	0.02	0.02
Rad	4.89	119.05	0.02	0.02
Rad	4.60	116.48	0.01	0.01
M	4.42	199.16	0.04	0.04
M	4.30	236.48	0.02	0.02
Rad	4.15			
Rad	3.93	133.62	0.02	0.02
Rad	3.71	136.75	0.02	0.02
Rad	3.51	136.82	0.02	0.02

CN-AA (Turbid gelator)

Axis	d-Spacings (Å)	Intensity	+d (Å)	-d (Å)
Rad (E)	13.35	122.58	0.20	0.19
Rad (E)	11.05	119.05	0.03	0.03
Rad (E)	9.25	130.82	0.14	0.14
Rad (E)	8.00	124.93	0.05	0.05
Rad (E)	6.35	169.87	0.06	0.06
Rad (E)	5.71	171.19	0.05	0.05
Rad (M)	4.57	214.20	0.01	0.01
Rad (M)	4.10	204.43	0.06	0.05
Rad (M)	3.55	178.22	0.01	0.01

2-AV

Axis	d-Spacings (Å)	Intensity	+d (Å)	-d (Å)
E	14.37	185.00	0.14	0.14
E	6.41	140.69	0.05	0.05
E	5.55	139.61	0.07	0.07
M	4.70	191.43	0.02	0.02

Br-AV

Axis	d-Spacings (Å)	Intensity	+d (Å)	-d (Å)
E	65.27	252.00	14.27	9.93
E	38.34	127.79	2.39	2.12
E	27.49	91.31	1.98	1.73
E	20.70	102.87	0.61	0.57
E	16.11	123.15	0.33	0.32
E	13.12	98.18	0.18	0.18
E	11.10	82.22	0.16	0.15
E	7.28	90.51	0.05	0.05
E	6.49	96.74	0.06	0.06
E	6.02	104.58	0.04	0.04
E	5.43	103.58	0.03	0.03
M	4.51	122.47	0.03	0.03
M	3.74	100.66	0.02	0.02
M	3.54	96.36	0.02	0.02
M	3.39	93.92	0.02	0.02

CN-AV

Axis	d-Spacings (Å)	Intensity	+d (Å)	-d (Å)
Rad (E)	14.05	127.57	0.08	-0.08
Rad (E)	12.61	130.42	0.02	-0.02
Rad (E)	10.67	140.48	0.01	-0.01
Rad (E)	9.42	161.57	0.02	-0.02
Rad (E)	7.39	131.09	0.05	-0.05
Rad (E)	6.25	239.86	0.00	0.00
Rad (M)	4.54	244.19	0.00	0.00
Rad (M)	4.16	228.97	0.03	-0.03

2-FF (D2O)

Axis	d-Spacings (Å)	Intensity	+d (Å)	-d (Å)
Rad (E)	28.75	124.00	2.47	2.11
Rad (E)	22.57	123.67	0.39	0.37
Rad (E)	16.45	104.81	0.23	0.24
Rad (E)	11.39	136.39	0.08	0.08
Rad (E)	8.11	106.68	0.14	0.14
Rad (M)	4.67	217.75	0.02	0.02

Br-FF

Axis	d-Spacings (Å)	Intensity	+d (Å)	-d (Å)
E	26.43	79.48	0.34	0.33
E	17.49	96.68	0.00	0.00
E	11.94	98.91	0.02	-0.02
E	8.21	94.86	0.01	0.01
E	6.00*			
E	5.45*			
M	4.71	197.03	0.02	-0.02
Rad	4.39	197.44	0.00	0.00
Rad	3.47	112.55	0.00	0.00

CN-FF

Axis	d-Spacings (Å)	Intensity	+d (Å)	-d (Å)
Rad (E)	23.39	158.63	1.59	1.40
Rad (E)	11.09	141.54	0.27	0.26
Rad (E)	7.75	109.54	0.01	0.01
Rad (E)	5.88*			
Rad (M)	4.66	214.88	0.06	0.06
Rad (M)	4.41	222.61	0.04	0.04
Rad (M)	4.11	190.03	0.00	0.00
Rad	3.42*			

2-FV

Axis	d-Spacings (Å)	Intensity	+d (Å)	-d (Å)
E	11.32	104.66	0.37	0.35
Rad (M)	4.70	195.74	0.02	0.02

Br-FV (Turbid gelator)

Axis	d-Spacings (Å)	Intensity	+d (Å)	-d (Å)
Rad (E)	23.37	141.89	1.01	0.93
Rad (E)	12.26	104.14	0.14	0.14
Rad (E)	10.06	99.13	0.09	0.09
Rad (E)	8.93	105.98	0.09	0.09
Rad (E)	7.53	147.16	0.05	0.05
Rad (E)	7.12	147.94	0.06	0.06
Rad (E)	6.00	165.23	0.04	0.04
Rad (E)	5.21	187.59	0.03	0.03
Rad (M)	4.66	211.84	0.03	0.03
Rad (M)	4.38	223.02	0.01	0.01
Rad (M)	4.17	207.58	0.02	0.02
Rad (M)	4.07	204.49	0.01	0.01
Rad (M)	3.91	183.76	0.01	0.01
Rad (M)	3.70	186.56	0.01	0.01
Rad (M)	3.57	201.10	0.01	0.01
Rad (M)	3.47	207.98	0.01	0.01
Rad (M)	3.36	188.92	0.01	0.01
Rad (M)	3.16	130.58	0.01	0.01
Rad (M)	3.06	123.15	0.00	0.00

CN-FV

Axis	d-Spacings (Å)	Intensity	+d (Å)	-d (Å)
Rad (E)	11.98	171.97	0.11	0.11
Rad (E)	10.46	151.07	0.01	0.01
Rad (E)	7.81	174.81	0.03	0.03
Rad (E)	6.46	171.44	0.01	0.01
Rad (E)	6.12	172.66	0.06	0.06
Rad (M)	4.64	209.44	0.00	0.00
Rad	4.46	203.82	0.05	0.05
Rad	4.14	203.62	0.03	0.03
Rad	3.46	162.48	0.00	0.00

Naphthalene dipeptide library unit cell searches

a

Cell dimension:	*a*	*b*	*c*	*a*	*b*	*c*	*a*	*b*	*c*	*a*	*b*	*c*	*a*	*b*	*c*
Cell search size:	**70**	**50**	**-**	**50**	**50**	**-**	**50**	**20**	**-**	**20**	**20**	**-**	**20**	**10**	**-**
2-AA	57.25	55.93	4.68	41.97	39.91	4.68	39.91	14.72	4.68						
2-AV	61.61	47.15	4.70	45.04	38.79	4.70	35.39	15.24	4.70	20.13	19.20	4.70	12.86	5.50	4.70
2-FF	60.39	32.00	4.67	44.66	34.44	4.67	40.26	18.95	4.67						
2-FV															
Br-AA	65.05	33.00	4.30	65.05	33.00	4.30	32.76	13.20	4.30						
Br-AV	65.66	48.20	4.51	38.95	26.97	4.51									
Br-FF	66.00	28.34	4.71	40.45	35.81	4.71	24.66	24.48	4.71						
Br-FV	68.62	42.64	4.66	45.84	42.48	4.66	44.74	14.19	4.66						
CN-AA	65.09	35.45	4.57	47.07	45.77	4.57	39.89	20.25	4.57	13.66	18.81	4.57			
CN-AV	64.14	37.67	4.54	41.82	37.60	4.54	46.93	14.78	4.54	13.01	14.56	4.54			
CN-FF	66.02	32.11	4.54	85.87	23.82	4.54	46.84	12.49	4.54						
CN-FV	64.97	17.78	4.64	35.91	25.97	4.64	46.49	14.32	4.64	12.23	19.63	4.64			
STD	3.16	10.63	0.12	14.52	7.02	0.12	7.21	3.77	0.12	3.63	2.35	0.07			

b

System	*a*	*b*	*c*	System	*a*	*b*	*c*	System	*a*	*b*	*c*
2-AA	39.91	14.72	4.68	**Br-AA**	32.76	13.20	4.30	**CN-AA**	39.89	20.25	4.57
2-AV	35.39	15.24	4.70	**Br-AV**				**CN-AV**	46.93	14.78	4.54
2-FF	-	-	-	**Br-FF**	24.66	24.48	4.71	**CN-FF**	46.84	12.49	4.54
2-FV	?	?	?	**Br-FV**	44.74	14.19	4.66	**CN-FV**	46.49	14.32	4.64
STD	3.20	0.37	0.01	**STD**	10.10	6.25	0.22	**STD**	3.44	3.34	0.05

c

System	*a*	*b*	*c*
Br-AA	65.05	33.00	4.30
Br-AV	65.66	48.20	4.51
Br-FF	66.00	28.34	4.71
Br-FV	68.62	42.64	4.66
STD	1.57	9.02	0.18

d

Grouped by naphthalene conjugate

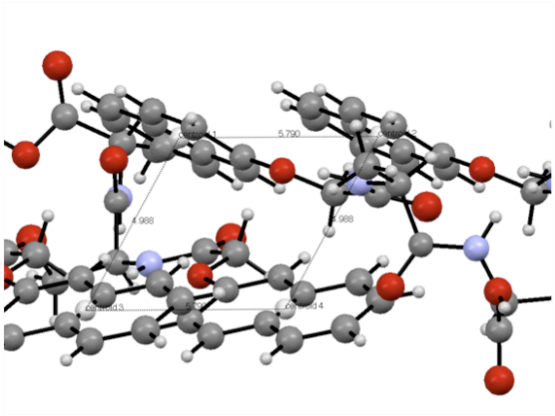
Cell searches a50, b20, cx					Cell searches a50, b20, cx. Including BrFV a70, b50, cx				
System	***a***	***b***	***c***	***Error***	***System***	***a***	***b***	***c***	***Error***
Nap	37.65	14.98	4.69		Nap	38.52	16.30	4.68	
±	3.20	0.37	0.01	1.68	±	2.72	2.31	0.02	1.68
Br-Nap	34.05	17.29	4.56		Br-Nap	66.33	38.04	4.55	
±	10.10	6.25	0.22	5.52	±	1.57	9.02	0.18	3.59
CN-Nap	45.04	15.46	4.57		CN-Nap	45.04	15.46	4.57	
±	3.44	3.34	0.05	2.28	±	3.44	3.34	0.05	2.28
			Grouped error:	3.16				**Grouped error:**	2.52

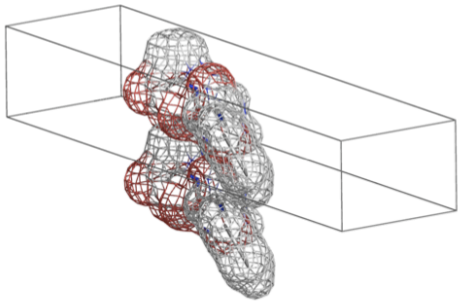
Grouped by sequence

Cell searches a50, b20, cx					Cell searches a50, b20, cx. Including BrFV a70, b50, cx				
System	***a***	***b***	***c***	***Error***	***System***	***a***	***b***	***c***	***Error***
AA	37.52	16.06	4.52		AA	48.28	22.66	4.52	
±	4.12	3.71	0.20	2.68	±	14.52	9.37	0.20	8.03
AV	41.16	15.01	4.62		AV	49.33	26.07	4.58	
±	8.16	0.33	0.11	2.87	±	15.28	19.16	0.10	11.51
FF	37.25	18.64	4.64		FF	51.03	19.93	4.64	
±	11.39	6.00	0.09	5.83	±	13.37	7.97	0.09	7.14
FV	45.61	14.26	4.65		FV	57.55	28.48	4.65	
±	1.24	0.09	0.01	0.45	±	15.65	20.02	0.01	11.90
			Grouped error:	2.95				**Grouped error:**	9.65

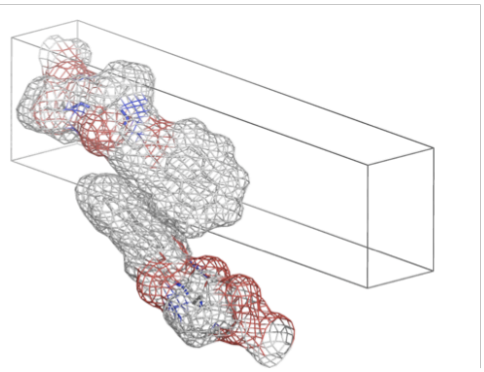
a) The lowest error unit cells determined for each dipeptide system over a range of initial unit cell search limits. Unit cells for 2-FV were not explored where only one equatorial reflection was observed. The a and b dimensions pertain to equatorial reflections and the c dimensions to meridional spacings.
b) The lowest error unit cells determined for each dipeptide system using a search limit of a 50, b 20, c x.
c) The lowest error unit cells determined for 6-bromo-2-naphthalene dipeptide systems using search limits of a 70 – 50, b 50 and c x.
d) The highest correlation in unit cell dimensions is found when grouping the predicted unit cells by naphthalene conjugate and including the large cell searches for Br-Nap dipeptides.

2AA crystal naphthalene centroid to centroid distances





Aromatic stacking along the *a* dimension at 5.79 Å

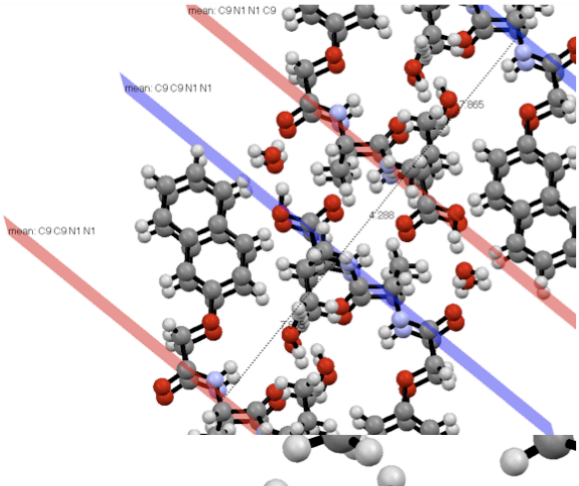


Aromatic stacking along the *ab* plane at 4.99 Å

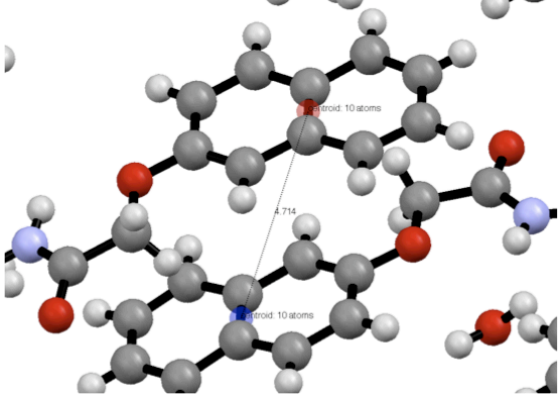
2AV crystal structure ß-strand separation and naphthalene centroid to centroid distance

ß-strand separation measured in Mercury software by drawing parallel planes through the first carbonyl carbon and nitrogen of the second amide bond.

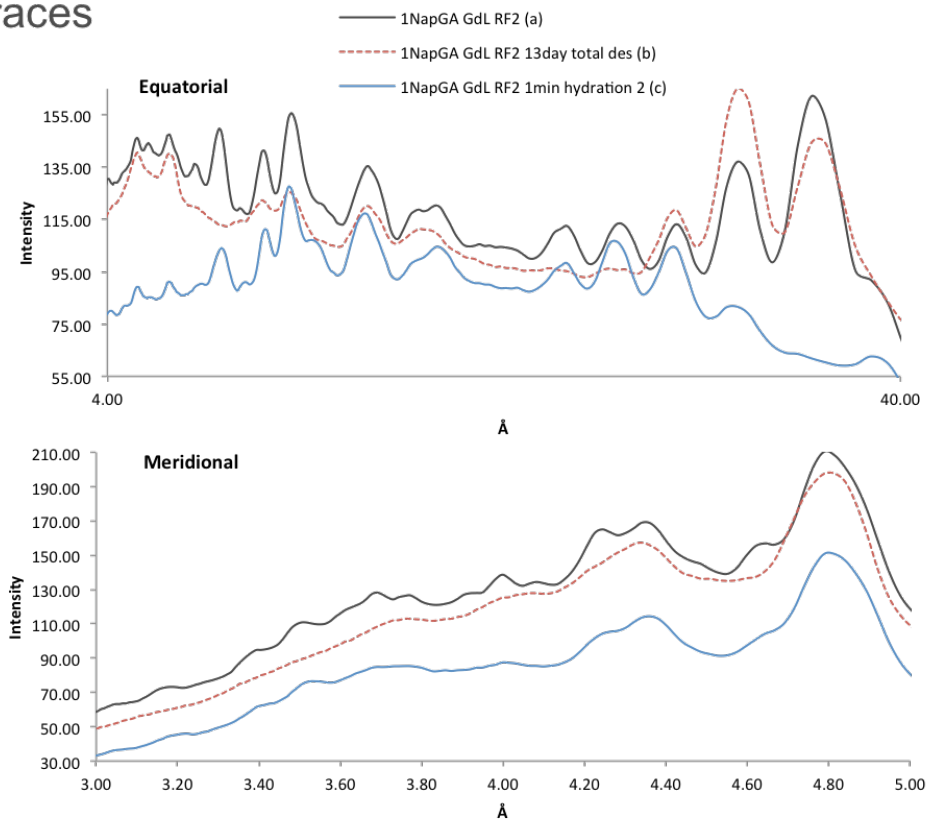
The dry interface is separated by 4.288 Å and the wet by 7.865.



The centroid to centroid distance of the naphthalene groups was measured to by 4.714 Å.



1NapGA fibre diffraction table of reflections and 1D traces



Pre-dehydration

Equatorials

d-Spacings (Å)	Intensity	+d (Å)	-d (Å)
32.41	164.80	0.24	0.24
25.87	139.54	0.08	0.08
21.45	115.16	0.07	0.07
18.09	114.77	0.05	0.05
15.39	113.53	0.03	0.03
10.49	122.00	0.01	0.01
8.57	137.24	0.01	0.01
6.84	157.42	0.01	0.01
6.29	142.03	0.01	0.01
5.52	150.61	0.01	0.01

Merdionals

d-Spacing (Å)	Intensity	+d (Å)	-d (Å)
4.76	193.35	0.02	0.02
4.60	153.54	0.02	0.02
4.30	166.92	0.02	0.02
4.19	161.07	0.02	0.02
4.02	137.91	0.01	0.01
3.94	143.50	0.01	0.01
3.84*	-	-	-
3.69	131.61	0.02	0.02

Dehydrated

Equatorials

d-Spacings (Å)	Intensity	+d (Å)	-d (Å)
32.70	152.08	0.13	0.13
25.85	169.89	0.09	0.09
21.29	122.05	0.00	0.00
10.07	113.40	0.04	0.04
8.56	121.42	0.02	0.02
6.80	125.54	0.01	0.01
6.28	122.48	0.02	0.02

Merdionals

d-Spacing (Å)	Intensity	+d (Å)	-d (Å)
4.76	208.92	0.03	0.03
4.29	161.42	0.01	0.01

Rehydrated

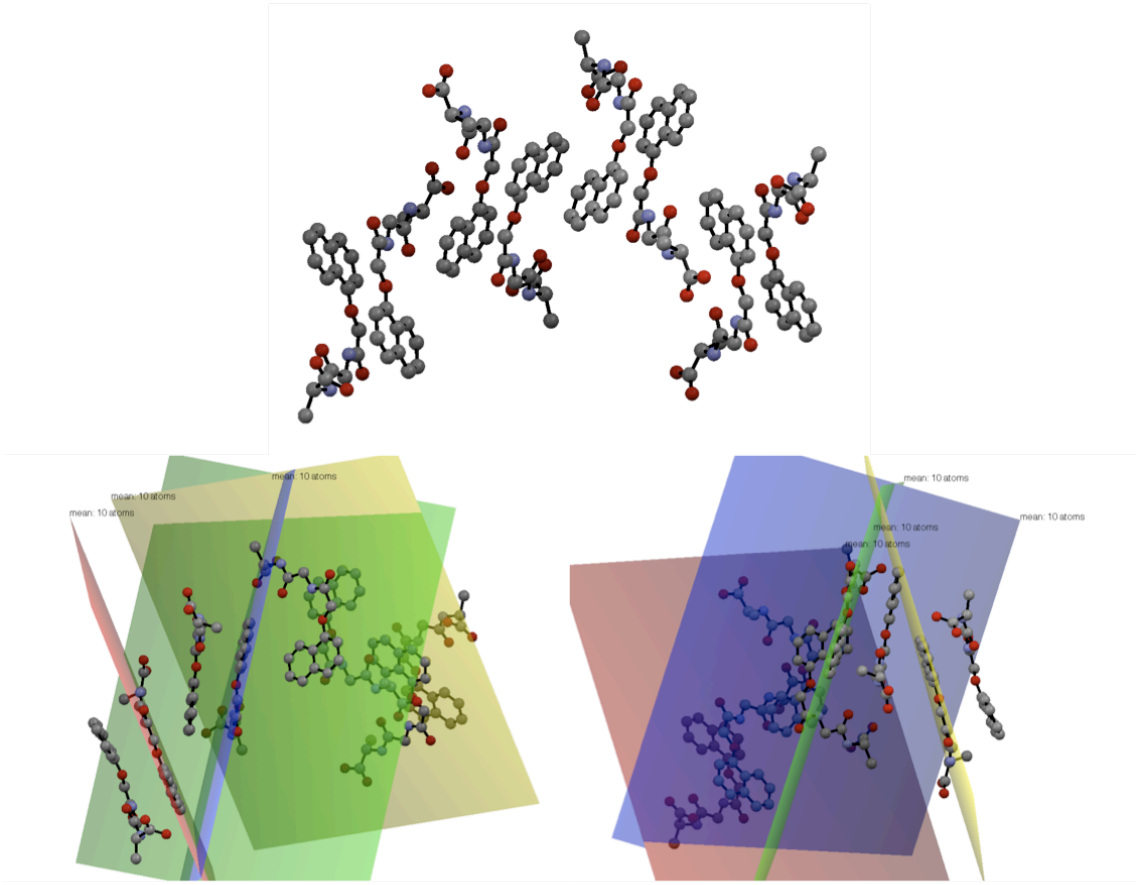
Equatorials

d-Spacings (Å)	Intensity	+d (Å)	-d (Å)
25.04	86.19	0.05	0.05
21.05	108.11	0.30	0.31
17.80	109.32	0.23	0.24
15.28	100.72	0.17	0.17
10.53	106.52	0.05	0.05
8.49	118.64	0.04	0.04
7.26*	-	-	-
6.81	126.17	0.03	0.03
6.33	110.95	0.01	0.01
5.90*	-	-	-
5.55	103.24	0.01	0.01
5.24	89.69	0.00	0.00

Merdionals

d-Spacing (Å)	Intensity	+d (Å)	-d (Å)
4.77	142.01	0.02	0.02
4.65*	-	-	-
4.30	112.99	0.02	0.02
4.23*	-	-	-

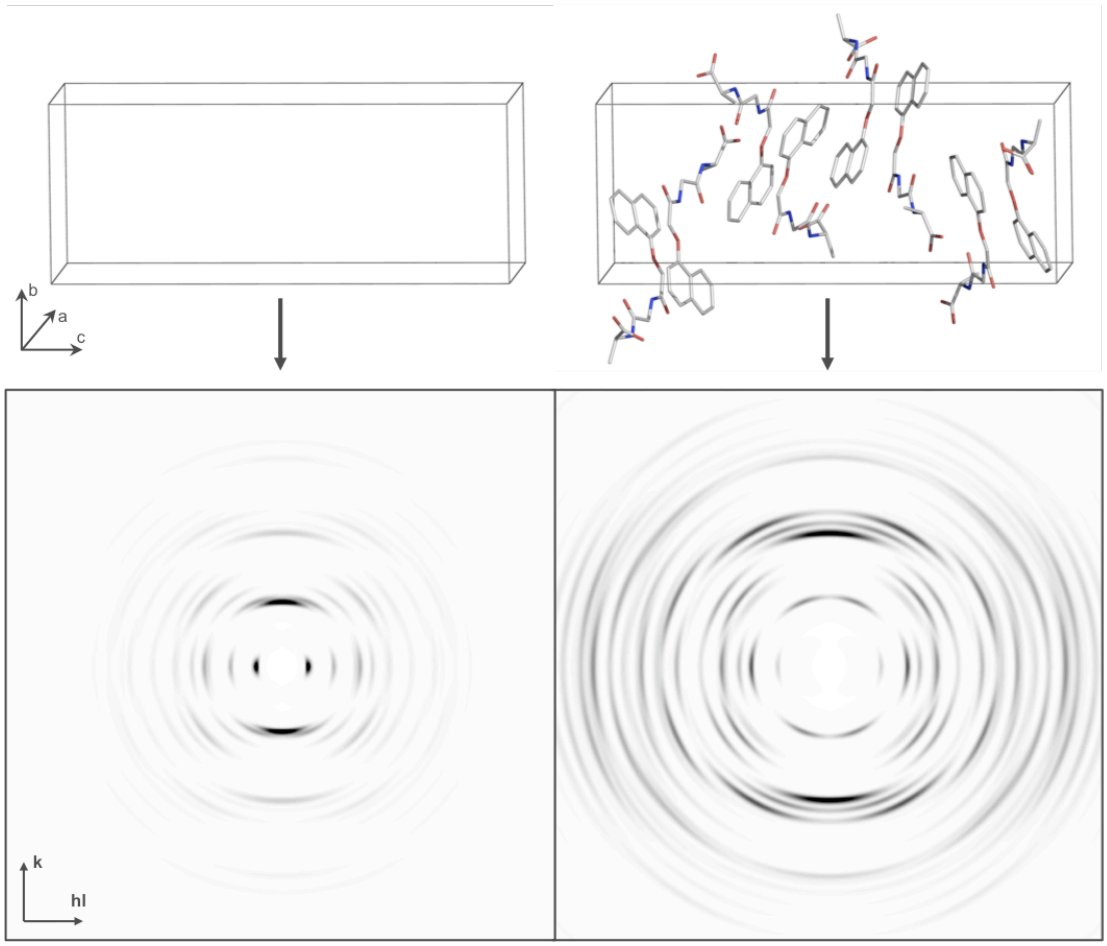
1NapGA crystal structure

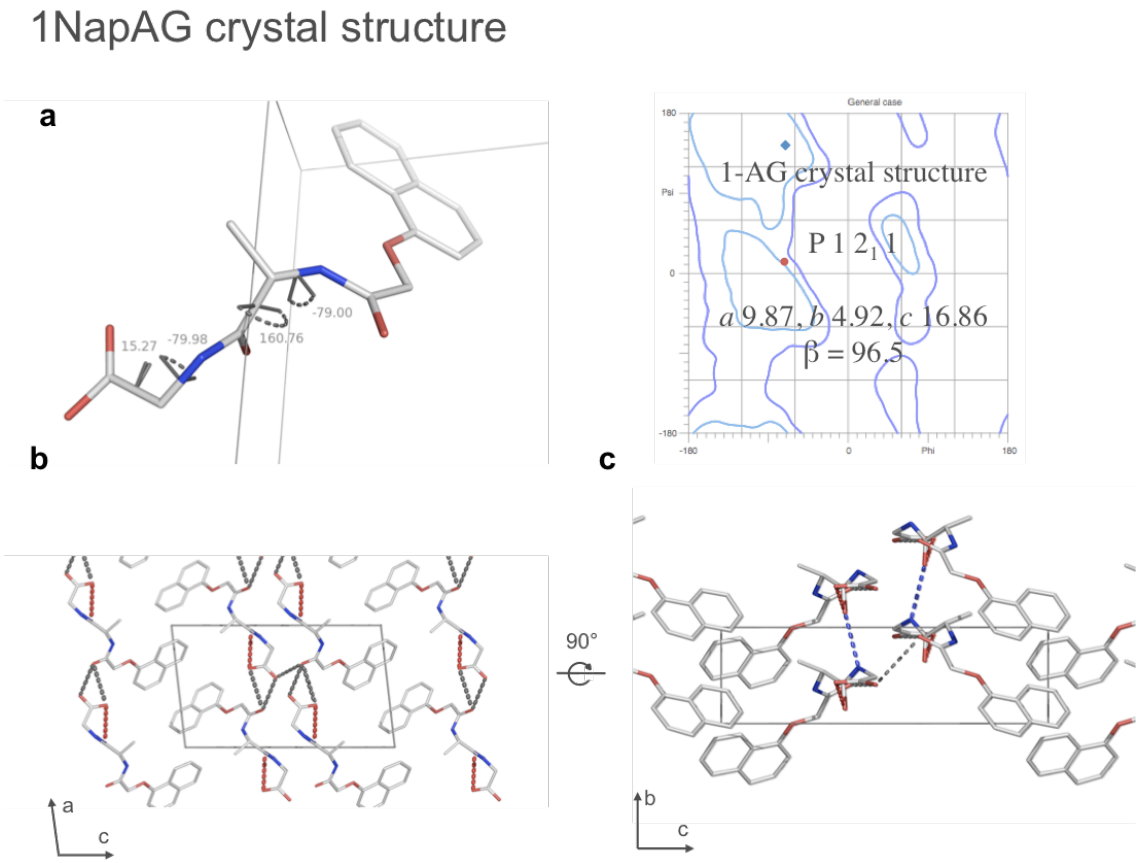


The 1GA unit cell (a), showing four structural planes viewed stereoscopically from a left (b) and right hand perspective (c). Graphics generated in Mercury.

1NapGA crystal structure structure and cell only simulation

a 9.92, b 11.11, c 28.25, α = β = γ = 90.0





The structure of 1AG in the crystalline phase. The conformation of the dipeptide is nonstandard with the second dihedral angle lying outside of Ramachandran allowed regions (a). The structure is stabilised by a two-dimensional hydrogen-bonding network involving intracarboxyl-carbonyl (red dashed lines), intercarboxyl-carbonyl (grey dashed lines) and interamide (blue dashed lines) interactions (b and c). Coordinates provided by Dr. Dave Adams (Liverpool University) and solved by Dr. Marc Schmidtmann (Liverpool University).

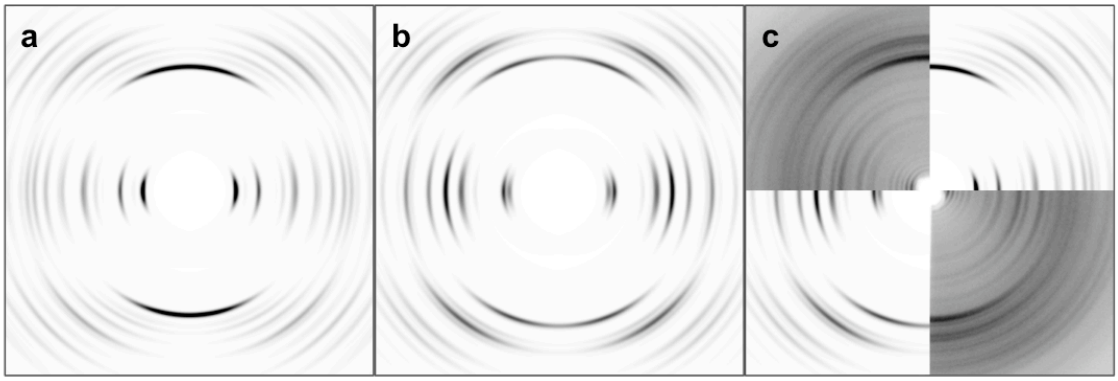
1NapGA structure prediction parameters

Provided by Dr. Graeme Day (Cambridge University)

Predictions parameters pertain to those found in (Adams, Morris et al. 2010)

Structure	Spacegroup	Total Energy ($kJmol^{-1}$)	Vol Per Molecule ($Å^3$)	a	b	c	α	β	γ
1	P 21	-178.31	416.17	13.47	4.28	14.43	90.00	89.77	90.00
2	P 21 21 21	-172.28	400.32	26.21	11.94	5.12	90.00	90.00	90.00
3	P 21	-171.83	403.95	12.87	5.03	13.52	90.00	112.65	90.00
4	P 21	-170.75	407.46	11.41	14.63	4.96	90.00	100.14	90.00
5	P 21	-169.84	417.53	11.11	4.90	15.71	90.00	77.53	90.00
6	P 21	-169.73	401.60	11.38	5.10	14.23	90.00	103.27	90.00
7	P 21	-169.35	408.94	4.93	15.08	12.44	90.00	117.88	90.00
8	P 21 21 21	-168.62	398.08	21.30	14.88	5.03	90.00	90.00	90.00
9	P 21	-168.50	420.07	13.25	4.31	14.73	90.00	89.62	90.00
10	P 21	-168.18	403.89	13.49	5.07	11.86	90.00	85.26	90.00
11	P 21	-168.12	418.36	13.19	4.97	13.55	90.00	109.71	90.00
12	P 21	-167.98	406.48	12.20	4.96	13.95	90.00	105.36	90.00
13	P 21	-167.17	411.82	16.35	4.92	10.92	90.00	110.25	90.00
14	P 21	-167.11	417.51	11.41	4.94	15.08	90.00	79.51	90.00
15	P 21 21 21	-167.00	398.01	14.84	5.20	20.62	90.00	90.00	90.00
16	P 21 21 21	-167.00	412.24	13.50	27.91	4.38	90.00	90.00	90.00
17	P 21	-166.94	406.15	8.05	6.90	16.30	90.00	116.17	90.00
18	P 21	-166.04	399.04	5.04	12.31	13.68	90.00	70.28	90.00
19	P 21	-165.97	413.20	11.34	5.05	14.63	90.00	80.89	90.00
20	P 21	-165.20	401.97	4.92	12.21	14.29	90.00	69.57	90.00
21	P 21	-164.95	424.63	11.38	4.99	15.65	90.00	72.84	90.00
22	P 21	-164.58	398.34	4.93	14.04	11.58	90.00	96.00	90.00
23	P 21	-164.56	407.02	14.75	5.93	9.38	90.00	96.62	90.00
24	P 21	-164.20	406.85	16.04	5.07	11.26	90.00	117.34	90.00
25	P 21	-164.07	409.55	15.33	5.04	10.93	90.00	103.97	90.00
26	P 21	-163.96	401.29	13.35	4.95	12.13	90.00	90.95	90.00
27	P 21	-163.28	417.41	12.02	5.01	14.78	90.00	69.80	90.00
28	P 21 21 21	-162.92	412.45	11.70	4.96	28.41	90.00	90.00	90.00
29	P 21	-162.73	413.22	13.07	4.91	12.91	90.00	86.26	90.00
30	P 21	-162.59	412.18	10.71	15.70	5.00	90.00	101.07	90.00
31	P 21	-162.52	409.85	12.57	7.59	8.94	90.00	105.94	90.00
32	P 21 21 21	-162.40	408.20	5.20	18.30	17.15	90.00	90.00	90.00
33	P 21	-162.35	403.84	10.97	4.94	15.63	90.00	107.62	90.00
34	P 21	-162.15	396.90	13.28	4.95	12.77	90.00	108.97	90.00
35	P 21	-161.73	419.01	5.04	14.79	11.25	90.00	88.93	90.00
36	P 21	-161.22	407.08	14.52	4.94	11.74	90.00	75.12	90.00
37	P 21	-160.99	411.69	10.52	5.08	15.91	90.00	75.60	90.00
38	P 21	-160.99	418.73	11.38	9.14	8.51	90.00	108.87	90.00
39	P 21 21 21	-160.93	417.58	8.12	13.31	15.45	90.00	90.00	90.00
40	P 21	-160.78	414.91	10.78	5.08	15.76	90.00	73.72	90.00
41	P 21	-160.77	414.07	15.09	4.91	11.44	90.00	102.59	90.00
42	P 21	-160.63	405.17	14.77	4.93	11.39	90.00	77.85	90.00
43	P 21	-160.63	409.78	4.91	11.16	15.29	90.00	101.82	90.00
44	P 21	-160.51	407.65	11.12	15.58	4.78	90.00	79.69	90.00
45	P 21	-160.40	418.07	12.65	8.27	8.40	90.00	72.03	90.00
46	P 21 21 21	-160.28	411.93	24.31	13.66	4.96	90.00	90.00	90.00
47	P 21 21 21	-159.98	415.02	25.08	13.15	5.03	90.00	90.00	90.00
48	P 21	-159.81	421.25	10.95	5.10	15.09	90.00	88.29	90.00
49	P 21 21 21	-159.50	402.79	5.05	12.70	25.13	90.00	90.00	90.00
50	P 21	-159.49	410.70	5.39	15.11	10.12	90.00	95.38	90.00
51	P 21 21 21	-159.46	426.19	23.89	9.15	7.80	90.00	90.00	90.00
52	P 21	-159.34	415.33	13.74	4.94	13.18	90.00	68.28	90.00
53	P 21	-159.10	412.76	11.65	5.14	13.90	90.00	97.49	90.00
54	P 21 21 21	-159.01	420.05	4.92	22.33	15.30	90.00	90.00	90.00
55	P 21	-158.94	422.25	12.15	14.16	5.02	90.00	78.12	90.00
56	P 21	-158.91	412.94	7.69	8.48	12.92	90.00	78.47	90.00
57	P 21	-158.84	414.00	7.31	20.17	6.12	90.00	66.68	90.00
58	P 21	-158.71	424.83	10.56	4.81	16.85	90.00	83.34	90.00
59	P 21	-158.71	402.69	13.37	12.12	4.97	90.00	90.77	90.00
60	P 21	-158.69	422.54	12.82	4.77	14.42	90.00	106.70	90.00
61	P 21	-158.67	407.53	11.68	14.37	4.97	90.00	102.36	90.00
62	P 21	-158.40	417.34	12.65	5.09	13.62	90.00	108.04	90.00
63	P 21	-158.29	416.74	13.27	4.92	13.12	90.00	103.16	90.00
64	P 21	-158.27	407.23	10.81	4.83	15.77	90.00	98.15	90.00
65	P 21 21 21	-158.27	429.83	12.94	9.28	14.31	90.00	90.00	90.00
66	P 21	-158.04	413.40	12.20	5.19	13.17	90.00	82.65	90.00
67	P 21 21 21	-157.69	417.78	9.47	16.72	10.56	90.00	90.00	90.00
68	P 21 21 21	-157.45	439.23	18.40	11.73	8.14	90.00	90.00	90.00
69	P 21	-157.41	406.19	9.87	4.94	17.13	90.00	76.47	90.00
70	P 21	-157.35	406.72	8.50	7.95	12.15	90.00	82.13	90.00
71	P 21	-157.28	412.07	4.97	9.99	16.80	90.00	98.61	90.00
72	P 21	-157.23	420.09	4.96	14.42	12.20	90.00	105.56	90.00
73	P 21	-157.16	403.00	17.32	5.07	9.27	90.00	82.49	90.00
74	P 21 21 21	-157.01	432.67	7.35	19.10	12.32	90.00	90.00	90.00
75	P 21 21 21	-156.93	410.11	5.00	27.46	11.95	90.00	90.00	90.00
76	P 21	-156.88	411.77	8.78	4.81	19.61	90.00	84.72	90.00
77	P 21	-156.55	405.31	10.60	4.90	17.70	90.00	118.14	90.00
78	P 21	-156.36	410.88	14.03	5.16	12.24	90.00	67.85	90.00
79	P 21	-156.21	417.71	13.36	4.94	12.86	90.00	100.29	90.00
80	P 21 21 21	-156.12	415.90	20.28	6.95	11.80	90.00	90.00	90.00
81	P 21	-156.07	404.04	11.69	14.78	5.03	90.00	111.41	90.00
82	P 21 21 21	-156.04	420.25	32.00	4.50	11.68	90.00	90.00	90.00
83	P 21	-155.96	407.31	16.38	4.85	11.31	90.00	64.98	90.00
84	P 21	-155.75	414.24	8.92	6.14	15.19	90.00	95.46	90.00
85	P 21 21 21	-155.68	398.24	30.71	9.35	5.55	90.00	90.00	90.00
86	P 21	-155.67	414.31	14.45	5.04	12.29	90.00	67.81	90.00
87	P 21	-155.66	405.99	11.05	5.10	14.58	90.00	98.71	90.00
88	P 21 21 21	-155.61	418.97	23.00	4.82	15.11	90.00	90.00	90.00
89	P 21 21 21	-155.55	425.99	9.02	13.19	14.32	90.00	90.00	90.00
90	P 21	-155.49	407.55	11.93	4.88	14.01	90.00	89.64	90.00
91	P 21 21 21	-155.36	410.82	7.31	9.27	24.25	90.00	90.00	90.00
92	P 21	-155.28	409.48	12.06	4.87	14.53	90.00	106.22	90.00
93	P 21 21 21	-155.24	410.15	6.89	9.37	25.40	90.00	90.00	90.00
94	P 21 21 21	-155.20	414.23	5.20	29.31	10.86	90.00	90.00	90.00
95	P 21	-155.00	412.21	11.19	4.88	15.82	90.00	107.53	90.00
96	P 21 21 21	-154.98	410.27	9.33	16.61	10.59	90.00	90.00	90.00
97	P 21	-154.89	423.33	10.85	4.93	15.88	90.00	85.25	90.00
98	P 21	-154.82	426.72	17.04	4.37	11.47	90.00	91.06	90.00
99	P 21 21 21	-154.81	410.55	5.04	20.79	15.68	90.00	90.00	90.00
100	P 21 21 21	-154.66	407.39	5.71	30.07	9.49	90.00	90.00	90.00

X-ray fibre diffraction simulation from 1GA packing modes



Simulated fibre diffraction from the two predicted packing modes of 1GA. The compact-column structure at -172.28 kJmol-1 in the orthorhombic cell a26.210, b11.942, c5.116 (a) and open-tape structure at -171.83 kJmol-1 in the monoclinic cell a12.869, b5.033, c13.516, beta112.65 (b) and their corresponding simulated patterns are shown assuming that the fibre axis is parallel to the interamide hydrogen bonding direction. Neither pattern can be successfully compared to the experimental fibre diffraction pattern of fibrillar 1NapGA (c). Structure oordinates provided by Dr. Dave Adams (Liverpool University) and determined by Dr. Graeme Day (Cambridge University).

Appendix IV

ImageJ (NIH) – Morphology picker script:

```
print("Morphology Picker v3.1");
print("Kyle Morris, University of Sussex, 2010");
print("--------------------");
		Dialog.create("Calibration, warning: global scales will be first removed");
		Dialog.addChoice("Magnification:", newArray("50,000X", "30,000X", "20,000X", "15,000X", "10,000X", "5,000X", "Manual"));
		Dialog.show();
		run("Set Scale...", "distance=0 known=0 pixel=1 unit=pixel global");
		Mag = Dialog.getChoice();
		print(Mag);
		if (Mag=="5,000X") {
			boxwidthpx = 37.5;
			boxlengthpx = 300;
		}
			else
			if (Mag=="10,000X") {
				boxwidthpx = 75;
				boxlengthpx = 600;
			}
				else
				if (Mag=="15,000X") {
					boxwidthpx = 112.5;
					boxlengthpx = 900;
				}
					else
					if (Mag=="20,000X") {
						boxwidthpx = 150;
						boxlengthpx = 1200;
					}
						else
						if (Mag=="30,000X") {
							boxwidthpx = 225;
							boxlengthpx = 1800;
						}
							else
							if (Mag=="50,000X") {
								boxwidthpx = 375;
								boxlengthpx = 3000;
							}
//Main function is embedded in while loop such that after each measurement user is asked if they want to repeat
		repeat = 1;
		do {
//Line parameters are collected
			waitForUser("Make a line selection and then click OK:");
			getLine(x1, y1, x2, y2, lineWidth);
			if (x1==-1) {
			exit("Can't find a line :*(, please restart macro");
			}
			getPixelSize(unit, pw, ph);
			x1*=pw; y1*=ph; x2*=pw; y2*=ph;
//Line parameters, x-y coordinates, length and angle are calculated
			dx = x2-x1; dy = y2-y1;
			length = sqrt(dx*dx+dy*dy);
			if (y1==y2) {
				if (x1<x2)
				angle = 0;
				else
				angle = 180;
			}
			else {
			angle = (atan2(y2-y1,x2-x1)*-57.2957795);
			}
			print("Selection parameters:");
			print ("Angle: ", angle, "degrees", "(", angle-90, ")");
			transformangle = (angle-90);
//User inputs fibril  in pixels and polygon is drawn at right angle to line
//User may adjust polygon width and length incrementally and when completed select the finalise checkbox to continue
			finalise=1;
			do {
				Dialog.create("");
				Dialog.addMessage("Enter fibril width:");
				Dialog.addNumber("Width/px:", boxwidthpx);
				Dialog.addNumber("Length/px:", boxlengthpx);
				Dialog.addCheckbox("Finalise:", false);
				Dialog.show();
				finalise = Dialog.getCheckbox();
				bwpx = Dialog.getNumber();
				blpx = Dialog.getNumber();;
				if (bwpx!=boxwidthpx||blpx!=boxlengthpx) {
					//print("px change");
					boxlengthpx = blpx;
					boxwidthpx = bwpx;
				}
				else {
					//print("No change");
				}
				lncX = (sin(angle/57.2957795)*boxwidthpx);
				if (angle==0) {
					lncY = boxwidthpx;
				}
				else {
					lncY = lncX/(tan(angle/57.2957795));
				}
				a = angle;
				b = 90 - angle;
				width = boxwidthpx/2;
				//print (width);
				length = boxlengthpx/2;
				//print (length);
				lncXnew = lncX/2;
				lncYnew = lncY/2;
				x1new = x1-(cos(b/57.2957795)*width);
				y1new = y1-(sin(b/57.2957795)*width);
				x2new = x1+(cos(angle/57.2957795)*boxlengthpx);
				y2new = y1-(sin(angle/57.2957795)*boxlengthpx);
				x2newer = x2new-(cos(b/57.2957795)*width);
				y2newer = y2new-(sin(b/57.2957795)*width);
				makePolygon(x1new,y1new,x2newer,y2newer,x2new+lncXnew,y2new+lncYnew,x1+lncXnew,y1+lncYnew);
				getStatistics(area, mean, min, max, std);
				background = mean;
```

```
                                        //print (background);
                                } while (finalise==0);
                                print("Length:", boxlengthpx, "pixels");
                                print("Width:", boxwidthpx, "pixels");
//Stores filename and position of extension for later renaming
name = getTitle();
//print(tmp);
ext = lastIndexOf(name, ".");
//print(ext);
//The following copies the selection made by the user and macro, copies to the internal clipboard and creates a new image with the copy rotated to normal.
run("Copy");
setBackgroundColor(background,0,0);
run("Internal Clipboard");
selectWindow("Clipboard");
run("Rotate... ", "angle=transformangle grid=1 interpolation=Bilinear fill enlarge");
//Updates new image filename with addition of "-CROP", minus extension
cropname = substring(name,0,ext);
cropname = cropname + "-CROP";
//print (cropname);
rename(cropname);
//Gets dimensions of new image
getDimensions(width, height, channels, slices, frames);
//print("x: ", width, "(", width/2, ")");
//print("y: ", height, "(", height/2, ")");
//Based on box selection calculates positions of polygon boudaries for new selection
x1 = width/2-boxwidthpx/2;
y1 = height/2-boxlengthpx/2;
x2 = width/2+boxwidthpx/2;
y2 = height/2-boxlengthpx/2;
x3 = width/2+boxwidthpx/2;
y3 = height/2+boxlengthpx/2;
x4 = width/2-boxwidthpx/2;
y4 = height/2+boxlengthpx/2;
makePolygon(x1, y1, x2, y2, x3, y3, x4, y4);
//Crops selection
run("Crop");
//Copies crop into new image, renames and runs bandpass filter with default settings
run("Copy");
run("Internal Clipboard");
selectWindow("Clipboard");
bpname = cropname + "-BP_FILTERED";
rename(bpname);
run("Bandpass Filter...", "filter_large=40 filter_small=3 suppress=None tolerance=5 autoscale saturate");
//Gives user the option of saving crop and bandpass filtered images to directory of original image
//Various overly complicated if statements figure out if the files are overwrites or not
selectWindow(name);
dir = getDirectory("image");
selectWindow(cropname);
selectWindow(bpname);
store = getBoolean("Do you want to save the cropped images to the parent directory?");
if (store==0) {
        print("--------------------");
        print("Warning: Processed images not saved");
        }
else {
        suffix = 1;
        Dialog.create("");
        Dialog.addNumber("Affix suffix to saved images?", suffix);
        Dialog.show();
        suffix = Dialog.getNumber();

        dir1 = dir + cropname + "-" + suffix;
        dir2 = dir + bpname + "-" + suffix;
        dir3 = dir + cropname + "-" + suffix + ".tif";
        dir4 = dir + bpname + "-" + suffix + ".tif";
        //print(dir);
        //print(dir1);
        //print(dir3);
        overwrite = File.exists(dir3);
        //print(overwrite);
        if (overwrite==1) {
                store1 = getBoolean("Suffix name exists. Overwrite?");
                //print("Overwrite=", store1);

                if (store1==0) {
                        print(store1);
                        print("--------------------");
                        print("Warning: Processed images not saved");
                        }
                else {
                        selectWindow(cropname);
                        saveAs("tiff", dir1);

                        selectWindow(bpname);
                        saveAs("tiff", dir2);

                        print("--------------------");
                        print("Processed images saved to parent directory:");
                        print(dir3);
                        print(dir4);
                }
                        }
        else {
                selectWindow(cropname);
                saveAs("tiff", dir1);

                selectWindow(bpname);
                saveAs("tiff", dir2);

                print("--------------------");
                print("Processed images saved to parent directory:");
                print(dir3);
                print(dir4);
        }
}
print("--------------------");
print("END OF MACRO");
print("--------------------");
exit();
```

PyMol lattice construction script:

```
#Kyle Morris, University of Sussex, 2010
#The following script can be run in pymol, using the run 'script.py' command.
#It will duplicate the monomer you give it and build a crystal based on user
#defined parameters (below).
#The starting monomer may have a custom prefix i.e. obj or P but it must be
#numbered 1 in its suffix i.e. obj1 or P1. Duplicated monomers are named sequentially.

######## INPUT CELL DIMENSIONS, SUPER CELL INDICES AND PREFIX HERE ##########

#Cell dimensions
a = 0
b = 5
c = 4.76

#Super cell indices
h = 0
k = 1
l = 40

#Prefix used
prefix = "obj"

###################### DO NOT ADJUST BEYOND THIS POINT ######################

#Total monomers required
end = h*k*l

#Starting monomer suffix number
start = 1

#Suffix number
suffix = 1

#Sets up a loop for h index
for i in range (0,h+1):
        hi = a*i

        #Sets up a loop for k index
        for j in range (0,k+1):
                kj = b*j

                #Sets up a loop for l index
                for m in range (0,l+1):
                        lm = c*m

                        source = str(prefix) + str(start)
                        copy = str(prefix) + str(suffix)

                        cmd.create(copy,source)
                        cmd.translate([hi,kj,lm], copy)

                        suffix = suffix + 1
```

MATLAB (MathWorks) – R-Factor pattern comparison function:

```
function XRFD_compare_exp_sim
%   MATLAB code for comparing an XRFD pattern to a CLEARER produced simulation
%   of equivalent format and size and saving difference map file to MATLAB
%   directory as ~R-factor image.tiff
%   The two images to be compared must be found in a directory that MATLAB has its
%   path pointing to.
%
%   AREAS THAT APPEAR BLACK ARE AREAS WHERE PIXEL INTENSITY IS IDENTICAL
%   WHITE AREAS SHOW DIFFERENCE :)

%User input for file names
filename1 = uigetfile('.tiff', 'Select experimental XRFD');
filename2 = uigetfile('.tiff', 'Select simulated XRFD');

%Image 1 matrix read in and conversion to gray scale
exp = imread(filename1);
im2double(exp);
%double(exp);
expD = ans;
expG = rgb2gray(expD);
expsize = size(expG);
%imshow(expG);

%Image 2 matrix read in and conversion to gray scale
sim = imread(filename2);
im2double(sim);
%double(sim);
simD = ans;
simsize = size(simD);

%simG = rgb2gray(simD);
%imshow(simG);

%Subtraction of image 1 & 2 matrices and conversion to absolute values
%(i.e. removal of sign, such that only absolute difference is measured)
comp = expG - simD;
comparison=abs(comp);

%Crop and mask setup
comparisoncrop = imcrop(comparison);
%%imshow(comparisoncrop);
mask = roipoly(comparisoncrop);
mask = ~mask;
masksize = sum(mask == 0);
masksizepx = sum(masksize)
%%imshow(mask);

%The mask matrix (mask values = 0) is multiplied by the comparison image
%matrix such that the masked area overlays the comparison image.
CCmask = comparisoncrop .* mask;

%The masked comparison is visualised and the sum of all the pixels
%calculated. Values of 1 indicate difference, values of 0 indicate no
%difference. The R-factor is calculated as the average of all the pixel
%difference values excluding the masked values.
imshow(CCmask);
RF1 = sum(abs(CCmask));
RF2 = sum(RF1);
S1 = size(CCmask);
x = [S1];
y = x(1)*x(2);
maskpcent = masksizepx/y*100;
RFactor = RF2/(y-masksizepx)

%The comparison image with and without mask is saved as .tiff files as well
%as the mask area itself
%A .txt file is created summarising the analysis
RFname = ['~R-Factor image RF ' num2str(RFactor) '.tiff'];
imwrite(comparisoncrop,RFname);
imwrite(CCmask, '~R-Factor image mask.tiff');
imwrite(mask,'~R-Factor image mask region.tiff');

msgbox(['R-Factor: ' num2str(RFactor)] ,'R-Factor')
log = fopen('~R-Factor log.txt', 'w');
fprintf(log, ['R-Factor comparison between: ' '\n\nImage #1: ' filename1 '\nPx size: ' num2str(expsize(1)) ' x ' num2str(expsize(2)) '\n\nImage #2: ' filename2 '\npx size: ' num2str(simsize(1)) ' x '
num2str(simsize(2)) '\n\nImage cropped to: ' num2str(x(1)) ' x ' num2str(x(2)) '\nTotal px size: ' num2str(y) '\n\nMask px size: ' num2str(masksizepx) ' (' num2str(maskpcent) ' Pcent) \n\nR-
Factor: ' num2str(RFactor)]);


end
```

Langmuir
Article
pubs.acs.org/Langmuir

Self-Assembly Mechanism for a Naphthalene−Dipeptide Leading to Hydrogelation

Lin Chen,[†] Kyle Morris,[‡] Andrea Laybourn,[†] David Elias,[†] Matthew R. Hicks,[§] Alison Rodger,[§] Louise Serpell,[‡] and Dave J. Adams*,[†]

[†]*Department of Chemistry, University of Liverpool, Crown Street, Liverpool, L69 7ZD, U.K.,* [‡]*Department of Chemistry and Biochemistry, School of Life Sciences, University of Sussex, Falmer, BN19QG, U.K., and* [§]*Department of Chemistry, University of Warwick, Coventry, CV4 7AL, U.K.*

Received September 30, 2009. Revised Manuscript Received November 5, 2009

Suitably functionalized dipeptides have been shown to be effective hydrogelators. The design of the hydrogelators and the mechanism by which hydrogelation occurs are both currently not well understood. Here, we have utilized the hydrolysis of glucono-δ-lactone to gluconic acid as a means of adjusting the pH in a naphthalene−alanylvaline solution allowing the specific targeting of the final pH. In addition, this method allows the assembly process to be characterized. We show that assembly begins as charge is removed from the *C*-terminus of the dipeptide. The removal of charge allows lateral assembly of the molecules leading to π−π stacking (shown by CD) and β-sheet formation (as shown by IR and X-ray fiber diffraction). This leads to the formation of fibrous structures. Electron microscopy reveals that thin fibers form initially, with low persistence length. Lateral association then occurs to give bundles of fibers with higher persistence length. This results in the initially weak hydrogel becoming stronger with time. The final mechanical properties of the hydrogels are very similar irrespective of the amount of GdL added; rather, the time taken to achieving the final gel is determined by the GdL concentration. However, differences are observed between the networks under strain, implying that the kinetics of assembly do impart different final materials' properties. Overall, this study provides detailed understanding of the assembly process that leads to hydrogelation.

Introduction

There is currently significant interest in the use of oligopeptides and oligopeptide−conjugates as hydrogelators.[1−3] Specific examples include β-sheet forming oligopeptides,[4−7] α-helix forming oligopeptides,[8] PEO−peptide conjugates,[9] and recently di- and tripeptides conjugated to large aromatic groups such as fluorenylmethoxycarbonyl (Fmoc)[10−16] or naphthalene.[17−19] Low molecular weight gelators (LMWG) such as these oligopeptides self-assemble in solution to form long fibers that trap the water, forming a gel.[20,21] As such, hydrogels prepared from LMWG differ from hydrogels prepared by the cross-linking of polymeric chains in that they are reversible since they are held together by noncovalent interactions. Another potential advantage is that LMWG often have stimuli-responsive reversible sol−gel transitions, for example by pH[10,11,22] or temperature cycles.[23] As a result, hydrogels prepared from LMWGs have been suggested as an attractive method for the preparation of novel materials for applications such as drug release[18] and tissue engineering.[8,11,12,24]

The majority of dipeptides conjugated to Fmoc or naphthalene groups form hydrogels at low pH (<4).[11,17] A rare example is Fmoc−diphenylalanine, which forms hydrogels at physiological pH.[25] Recently, it has been shown that this is due to dramatic pK_a shifts induced by the high hydrophobicity of the molecule, with the assembly being controlled by the charge on the dipeptide.[26] However, it is not yet understood whether these observations are general or highly specific to a single system. Additionally, it is clear that the kinetics of the pH adjustment process are key in governing the final materials' properties of the hydrogels formed. Values for the storage modulus (G') for hydrogels prepared from Fmoc−diphenylalanine vary from 1 to 10^4 Pa.[14,25,26] We are interested in understanding the link between the kinetics of assembly and a material's final properties (which are vital for any specific application).

We recently demonstrated that the assembly process of a range of Fmoc−dipeptides could be followed utilizing the hydrolysis of glucono-δ-lactone (GdL) to adjust the pH.[10] GdL hydrolyzes slowly in water,[27] forming gluconic acid and hence lowering the pH of a solution in a controlled manner. To our knowledge, there is not a comparable system that can be added to a pH-dependent process to provide a controlled adjustment of pH. Here, gelation is pH-dependent, and so our aim was to use this reaction to provide the required pH change uniformly across the solution. The slow GdL hydrolysis (as compared to fast dissolution) allows the gelation process to be followed with time. In addition, the final pH achieved can be targeted by adjusting the amount of GdL added. Hence, this method can be used to prepare gels quiescently in the absence of a shear or processing history that often dominates the outcome of the gelation process. As noted above, for many systems, the outcome of the pH change is history-dependent and therefore different even between samples that are nominally identical. Using this methodology, we report here on the assembly of a naphthalene−dipeptide, utilizing the adjustable kinetics of pH change to follow the assembly process. We show that the assembly process translates directly onto the mechanical properties of the hydrogel, allowing a detailed understanding of the hydrogelation process.

Experimental Section

Materials. All chemicals and solvents were purchased from Sigma-Aldrich and used as received. Millipore water (resistivity = 18.2 MΩ·cm) was used throughout.

Naphthalene−Dipeptide Synthesis. The peptide derivative was prepared from 6-bromo-2-naphthol as follows.

To a stirred solution of 6-bromo-2-naphthol (5.0 g, 22.4 mmol) and potassium carbonate (15.48 g, 5 equiv, 112 mmol) in acetone (130 mL) was added chloro-*tert*-butyl acetate (3.52 mL, 1.1 equiv, 24.64 mmol). The solution was heated to reflux overnight. After this time, chloroform (100 mL) was added, and the solution was washed with water (4 × 100 mL). The organic phase was dried with magnesium sulfate and the solvent removed *in vacuo*. The crude product was purified by flash column chromatography, eluting with 10% ethyl acetate in hexane, to give *tert*-butyl 2-(6-bromonaphthalen-2-yloxy)acetate as an off-white solid in a 79% yield. ^{1}H NMR ($CDCl_3$): 7.92 (d, ArH, 1H, $^3J_{HH}$ = 2.0 Hz), 7.67 (d, ArH, 1H, $^3J_{HH}$ = 9.0 Hz), 7.57 (d, ArH, 1H, $^3J_{HH}$ = 8.8 Hz), 7.50 (dd, ArH, 1H, $^3J_{HH}$ = 2.0 Hz), 7.24 (dd, ArH, 1H, $^3J_{HH}$ = 2.6 Hz), 7.02 (d, ArH, 1H, $^3J_{HH}$ = 2.6 Hz), 4.62 (s, OCH_2C═O, 2H), 1.50 ppm (s, $OC(CH_3)_3$, 9H). ^{13}C NMR ($CDCl_3$): 168.2, 156.5, 133.2, 130.8, 130.1 (d, J = 7.2 Hz), 129.2, 128.9, 120.1, 117.9, 107.5, 83.0, 66.2, 28.5, 28.3 ppm. MS (CI, NH_3) 354 ([M + NH_4]$^+$). Accurate mass calculated: 354.070 48. Found: 354.069 75. Analysis calculated for $C_{16}H_{17}O_3Br$: C, 56.99%; H, 5.08%. Found: C, 56.90%; H, 5.06%.

The *tert*-butyl protecting group was removed by dissolving *tert*-butyl 2-(6-bromonaphthalen-2-yloxy)acetate (8.0 g, 23.7 mmol) in chloroform (ca. 30 mL). Trifluoroacetic acid (11.08 mL, 6.1 equiv, 145 mmol) was added, and the solution was stirred for 24 h. Hexane (200 mL) was added, and the resulting white precipitate was collected. After washing with hexane, the resulting solid was dried *in vacuo* to give 2-(6-bromonaphthalen-2-yloxy)acetic acid in a 92% yield. ^{1}H NMR (d_6-DMSO): 8.13 (d, ArH, 1H, $^3J_{HH}$ = 1.5 Hz), 7.85 (d, ArH, 1H, $^3J_{HH}$ = 9.0 Hz), 7.78 (d, ArH, 1H, $^3J_{HH}$ = 8.8 Hz), 7.57 (dd, ArH, 1H, $^3J_{HH}$ = 2.2 Hz), 7.32 (d, ArH, 1H, $^3J_{HH}$ = 2.6 Hz), 7.26 (dd, ArH, 1H, $^3J_{HH}$ = 2.6 Hz), 4.81 (s, OCH_2, 2H), 3.39 ppm (s, O*H*). ^{13}C (d_6-DMSO): 170.6, 150.7, 134.9, 132.2, 128.7, 127.1, 126.2, 126.1, 124.3, 123.0, 122.5, 72.3, 59.9 ppm. MS (CI, NH_3) 298 ([M + NH_4]$^+$), Accurate mass calculated: 298.007 88. Found: 298.007 99. Analysis calculated for $C_{12}H_9O_3Br$: C, 51.27%; H, 3.23%. Found: C, 50.91%; H, 3.18%.

The C-ethyl-protected alanine was coupled to 2-(6-bromonaphthalen-2-yloxy)acetic acid using standard coupling methodologies.[28] To a stirred solution of 2-(6-bromonaphthalen-2-yloxy)acetic acid (1.00 g, 3.56 mmol) in chloroform (20 mL) was added *N*-methylmorpholine (0.40 mL, 3.60 mmol), and isobutyl chloroformate (0.49 mL, 3.60 mmol) at 0 °C. The solution was stirred at 0 °C for 5 min. A solution of alanine ethyl ester (0.35 g, 2.5 mmol) and *N*-methylmorpholine (0.40 mL, 3.60 mmol) in chloroform (20 mL) was added. The solution was allowed to warm to room temperature and stirring continued overnight. The solution was washed with distilled water (2 × 100 mL), hydrochloric acid (2 × 100 mL, 0.1 M), aqueous potassium carbonate solution (100 mL, 0.1 M), and distilled water again (4 × 100 mL) and dried over magnesium sulfate, and the solvent was removed *in vacuo*. The solid was then washed with petroleum ether (10 mL) and dried under vacuum for 48 h to give ethyl 2-(2(6-bromonaphthalen-2-yloxy)acetamido)propanoate in a 73% yield. ^{1}H NMR ($CDCl_3$): 7.94 (s, ArH, 1H), 7.70 (d, ArH, 1H, $^3J_{HH}$ = 8.8 Hz), 7.62 (d, ArH, 1H, $^3J_{HH}$ = 8.3 Hz), 7.53 (d, ArH, 1H, $^3J_{HH}$ = 8.3 Hz), 7.26 (m, ArH, 2H, $^3J_{HH}$ = 8.3 Hz), 7.12 (s, ArH, 1H), 4.71 (m, CH_3CHC═O, 1H), 4.62 (s, OCH_2C═O, 2H), 4.23 (m, OCH_2CH_3, 2H, $^3J_{HH}$ = 7.0 Hz), 1.49 (d, NHCH(CH_3), 3H, $^3J_{HH}$ = 6.6 Hz), 1.29 ppm (t, OCH_2CH_3, 3H, $^3J_{HH}$ = 7.0 Hz). ^{13}C NMR ($CDCl_3$): 172.9, 167.8, 155.7, 133.1, 130.9, 130.4, 130.1, 129.0, 119.7, 118.3, 108.0, 67.7, 62.1, 48.2, 18.9, 14.5 ppm. MS (ES+, CH_3OH) 402 ([M + Na]$^+$), Accurate mass calculated: 402.0317. Found: 402.0317. Analysis calculated for $C_{17}H_{18}NO_4Br$: C, 53.70%; H, 4.77%; N, 3.68%. Found: C, 54.04%; H, 4.78%; N, 3.66%.

To deprotect the *C*-terminus, lithium hydroxide (0.2 g) was added to a solution of ethyl 2-(2(6-bromonaphthalen-2-yloxy)acetamido)propanoate (0.5 g) in a THF:water solution (3:1 mixture, 25 mL), and the solution was stirred for 18 h. After this time, distilled water (ca. 100 mL) was added, and then hydrochloric acid (1.0 M) was added dropwise until the pH was lowered to pH 3. The resulting white precipitate was collected by filtration and washed well with water and petroleum ether before being dried *in vacuo* to give 2-(2-(6-bromonaphthalen-2-yloxy)acetamido)propanoic acid as a white powder in 89% yield. ^{1}H NMR (d_6-DMSO): 8.44 (d, ArH, 1H, $^3J_{HH}$ =7.5 Hz) 8.13 (d, ArH, 1H, $^3J_{HH}$ = 1.6 Hz), 7.86 (d, ArH, 1H, $^3J_{HH}$ = 8.8 Hz), 7.75 (d, ArH, 1H, $^3J_{HH}$ = 8.8 Hz), 7.58 (dd, ArH, 1H, $^3J_{HH}$ = 1.9 Hz), 7.34 (m, NH, 1H), 7.31 (dd, ArH, 1H, $^3J_{HH}$ = 2.4 Hz), 4.65 (m, OCH_2-C═O, 2H), 4.30 (m, $(CH_3)CHC$═O, 1H), 3.42 (s, OH), 1.33 (d, (CH_3)CHC═O, 3H, $^3J_{HH}$ = 7.2 Hz) ppm. ^{13}C (d_6-DMSO): 174.3, 167.5, 156.4, 133.0, 130.3, 129.7 (d, J = 5.7 Hz), 129.3, 129.0, 120.2, 116.9, 107.8, 67.1, 47.8, 17.6 ppm. MS (ES+, CH_3OH) 374 ([M + Na]$^+$). Accurate mass calculated: 374.0005. Found: 374.0004. Analysis calculated for $C_{15}H_{14}NO_4Br$: C, 51.16%; H, 4.01%; N, 3.96%. Found C, 51.16%; H, 4.02%; N, 3.96%.

A similar procedure was followed for the addition of the second amino acid.

Ethyl 2-(2-(2-(6-bromonaphthalen-2-yloxy)acetamido)propanamido)-3-methylbutanoate was prepared as a white solid in a 91% yield. ^{1}H NMR ($CDCl_3$): 7.92 (d, ArH, 1H, $^3J_{HH}$ = 1.6 Hz), 7.68 (d, ArH, 1H, $^3J_{HH}$ = 9.8 Hz), 7.59 (d, ArH, 1H, $^3J_{HH}$ = 8.7 Hz), 7.51 (dd, ArH, 1H, $^3J_{HH}$ = 8.7 Hz, $^3J_{HH}$ = 2.0 Hz), 7.22

*Corresponding author. E-mail: d.j.adams@liverpool.ac.uk.

(1) Ulijn, R. V.; Smith, A. M. *Chem. Soc. Rev.* **2008**, *37*, 664–675.
(2) Xu, B. *Langmuir* **2009**, *25*, 8375–8377.
(3) Bhattacharya, S.; Maitra, U.; Mukhopadhyay, S.; Srivastova, A. In *Molecular Gels: Materials with Self-Assembled Fibrillar Networks*; Weiss, R. G., Terech, P., Eds.; Springer: New York, 2005.
(4) Saiani, A.; Mohammed, A.; Frielinghaus, H.; Collins, R.; Hodson, N.; Kielty, C. M.; Sherratt, M. J.; Miller, A. F. *Soft Matter* **2009**, *5*, 193–202.
(5) Branco, M. C.; Pochan, D. J.; Wagner, N. J.; Schneider, J. P. *Biomaterials* **2009**, *30*, 1339–1347.
(6) Hule, R. A.; Nagarkar, R. P.; Altunbas, A.; Ramay, H. R.; Branco, M. C.; Schneider, J. P.; Pochan, D. J. *Faraday Discuss.* **2008**, *139*, 251–264.
(7) Krysmann, M. J.; Castelletto, V.; Kelarakis, A.; Hamley, I. W.; Hule, R. A.; Pochan, D. J. *Biochemistry* **2008**, *47*, 4597–4605.
(8) Banwell, E. F.; Abelardo, E. S.; Adams, D. J.; Birchall, M. A.; Corrigan, A.; Donald, A. M.; Kirkland, M.; Serpell, L. C.; Butler, M. F.; Woolfson, D. N. *Nat. Mater.* **2009**, *8*, 596–600.
(9) Tzokova, N.; Fernyhough, C. M.; Topham, P. D.; Sandon, N.; Adams, D. J.; Butler, M. F.; Armes, S. P.; Ryan, A. J. *Langmuir* **2009**, *25*, 2479–2485.
(10) Adams, D. J.; Butler, M. F.; Frith, W. F.; Kirkland, M.; Mullen, L.; Sanderson, P. *Soft Matter* **2009**, *5*, 1856–1862.
(11) Jayawarna, V.; Ali, M.; Jowitt, T. A.; Miller, A. E.; Saiani, A.; Gough, J. E.; Ulijn, R. V. *Adv. Mater.* **2006**, *18*, 611–614.
(12) Jayawarna, V.; Smith, A.; Gough, J. E.; Ulijn, R. V. *Biochem. Soc. Trans.* **2007**, *35*, 535–537.
(13) Williams, R. J.; Smith, A. M.; Collins, R.; Hodson, N.; Das, A. K.; Ulijn, R. V. *Nat. Nanotechnol.* **2009**, *4*, 19–24.
(14) Mahler, A.; Reches, M.; Rechter, M.; Cohen, S.; Gazit, E. *Adv. Mater.* **2006**, *18*, 1365–1370.
(15) Yang, Z. M.; Gu, H. W.; Fu, D. G.; Gao, P.; Lam, J. K.; Xu, B. *Adv. Mater.* **2004**, *16*, 1440–1444.
(16) Jayawarna, V.; Richardson, S. M.; Hirst, A. R.; Hodson, N. W.; Saiani, A.; Gough, J. E.; Ulijn, R. V. *Acta Biomaterialia* **2009**, *5*, 934–943.
(17) Yang, Z. M.; Liang, G. L.; Ma, M. L.; Gao, Y.; Xu, B. *J. Mater. Chem.* **2007**, *17*, 850–854.
(18) Liang, G. L.; Yang, Z. M.; Zhang, R. J.; Li, L. H.; Fan, Y. J.; Kuang, Y.; Gao, Y.; Wang, T.; Lu, W. W.; Xu, B. *Langmuir* **2009**, *25*, 8419–8422.
(19) Yang, Z. M.; Liang, G. L.; Xu, B. *Chem. Commun.* **2006**, 738–740.
(20) de Loos, M.; Feringa, B. L.; van Esch, J. H. *Eur. J. Org. Chem.* **2005**, 3615–3631.
(21) Estroff, L. A.; Hamilton, A. D. *Chem. Rev.* **2004**, *104*, 1201–1217.
(22) Xing, B. G.; Yu, C. W.; Chow, K. H.; Ho, P. L.; Fu, D. G.; Xu, B. *J. Am. Chem. Soc.* **2002**, *124*, 14846–14847.
(23) Vegners, R.; Sheshtakova, I.; Kalvinsh, I.; Ezzell, R. M.; Janmey, P. A. *J. Peptide Sci.* **1995**, *1*, 371–378.
(24) Zhou, M.; Smith, A. M.; Das, A. K.; Hodson, N. W.; Collins, R. F.; Ulijn, R. V.; Gough, J. E. *Biomaterials* **2009**, *30*, 2523–2530.
(25) Smith, A. M.; Williams, R. J.; Tang, C.; Coppo, P.; Collins, R. F.; Turner, M. L.; Saiani, A.; Ulijn, R. V. *Adv. Mater.* **2008**, *20*, 37–41.
(26) Tang, C.; Smith, A. M.; Collins, R. F.; Ulijn, R. V.; Saiani, A. *Langmuir* **2009**, *25*, 9447–9453.
(27) Pocker, Y.; Green, E. *J. Am. Chem. Soc.* **1973**, *95*, 113–119.
(28) Adams, D. J.; Young, I. *J. Polym. Sci., Part A: Polym. Chem.* **2008**, *46*, 6082–6090.

(dd, ArH, 1H, $^3J_{HH}$ = 9.1 Hz, $^3J_{HH}$ = 2.7 Hz), 7.16 (bd, NH, 1H, $^3J_{HH}$ = 7.3 Hz), 7.12 (d, ArH, 1H, $^3J_{HH}$ = 2.5 Hz), 6.52 (bd, NH, 1H, $^3J_{HH}$ = 8.8 Hz), 4.64 (m, C*H*NH, 1H), 4.51 (d, OCH_2, 2H, $^3J_{HH}$ = 2.2 Hz), 4.48 (d, C*H*NH, 1H, $^3J_{HH}$ = 8.7 Hz, $^3J_{HH}$ = 4.8 Hz), 4.21 (q, CH_2CH_3, 2H, $^3J_{HH}$ = 7.1 Hz), 2.15 (m, C*H*-$(CH_3)_2$, 1H), 1.44 (d, $CHCH_3$, 3H, $^3J_{HH}$ = 6.7 Hz), 1.28 (t, CH_2CH_3, 3H, $^3J_{HH}$ = 7.1 Hz), 0.91 (d, $CH(CH_3)$, 3H, $^3J_{HH}$ = 7.0 Hz), 0.89 ppm (d, $CH(CH_3)$, 3H, $^3J_{HH}$ = 7.0 Hz). ^{13}C NMR ($CDCl_3$): 171.5, 171.4, 167.7, 155.4, 132.8, 130.7, 130.0, 129.7, 129.0, 128.0, 119.2, 117.9, 107.9, 67.4, 61.2, 57.4, 48.6, 31.1, 18.8, 18.2, 17.6, 14.1 ppm. MS (CI) 498/496 ($[M + NH_4]^+$). Analysis calculated for $C_{22}H_{27}N_2O_5Br$: C, 55.12%; H, 5.68%; N, 5.84%. Found C, 55.03%; H, 5.72%; N, 5.81%.

2-(2-(2-(6-Bromonaphthalen-2-yloxy)acetamido)propanamido)-3-methylbutanoic acid was collected as a white solid in a 88% yield. ^{1}H NMR (d_6-DMSO): 8.20 (bd, NH, 1H, $^3J_{HH}$ = 7.6 Hz), 8.12 (d, ArH, 1H, $^3J_{HH}$ = 1.6 Hz), 8.06 (bd, NH, 1H, $^3J_{HH}$ = 8.5 Hz), 7.86 (d, ArH, 1H, $^3J_{HH}$ = 8.9 Hz), 7.76 (d, ArH, 1H, $^3J_{HH}$ = 8.8 Hz), 7.58 (dd, ArH, 1H, $^3J_{HH}$ = 8.8 Hz, $^3J_{HH}$ = 2.1 Hz), 7.34 (d, ArH, 1H, $^3J_{HH}$ = 1.6 Hz), 7.30 (dd, ArH, 1H, $^3J_{HH}$ = 8.9 Hz, $^3J_{HH}$ = 2.1 Hz), 4.66 (d, OCH_2, 2H, $^3J_{HH}$ = 2.2 Hz), 4.53 (m, C*H*NH, 1H), 4.15 (dd, C*H*NH, 1H, $^3J_{HH}$ = 8.6 Hz, $^3J_{HH}$ = 5.9 Hz), 2.04 (m, C*H*$(CH_3)_2$, 1H), 1.28 (d, $CHCH_3$, 3H, $^3J_{HH}$ = 7.1 Hz), 0.85 ppm (d, $CH(CH_3)_2$, 6H, $^3J_{HH}$ = 6.6 Hz). ^{13}C NMR (d_6-DMSO): 172.8, 172.2, 166.9, 155.9, 132.6, 129.9, 129.3, 129.2, 128.9, 128.6 119.7, 116.5, 107.5, 66.8, 57.1, 47.6, 29.8, 19.0, 18.4, 17.9 ppm. MS (CI) 468/470 ($[M + NH_4]^+$). Analysis calculated for $C_{20}H_{23}N_2O_5Br$: C, 53.23%; H, 5.14%; N, 6.21%. Found C, 52.93%; H, 5.11%; N, 6.18%.

NMR. ^{1}H NMR spectra were recorded at 400.13 MHz using a Bruker Avance 400 NMR spectrometer. ^{13}C NMR spectra were recorded at 100.6 MHz.

Gelation Studies. The dipeptide derivative (25.0 mg) was suspended in deionized water (5.00 mL). An equimolar quantity of NaOH (0.1 M, aq) was added, and the solution was gently stirred for 30 min until a clear solution was formed. The pH of this solution was measured to be 10.7. To prepare hydrogels, solutions were added to measured quantities of glucono-δ-lactone (GdL) according to the requirements of final pH, and the samples were left to stand for 24 h.

pH Measurements and pK_a Determination. A FC200 pH probe (HANNA instruments) with a (6 mm × 10 mm) conical tip was employed for all the pH measurements. The stated accuracy of the pH measurements is ±0.1. The pH changes during the gelation process were recorded every 1 min for 24 h. All measurements were conducted at room temperature. The pK_a values of dipeptide derivative (0.5 wt %) solutions were determined by titration via the addition of aliquots of a 0.1 M HCl solution. pH values were also recorded every 2 s during the titration process. To prevent the formation of gel, the solutions were gently stirred, thus keeping the solution liquid during the whole "titration" process.

Circular Dichroism and Linear Dichroism. Measurements were performed at room temperature (21 °C) on a Jasco J-815 spectropolarimeter modified for linear dichroism (Jasco UK, Great Dunmow, UK). Spectra were measured between 350 and 180 nm with a data pitch of 0.2 nm. The bandwidth was set to 2 nm with a scanning speed of 100 nm min^{-1} and a response time of 1 s. Spectra were collected at 2 min intervals. For clarity, the circular dichroism (CD) and linear dichroism (LD) figures show spectra only every 10 min. The path length was 0.1 mm in a demountable Spectrosil quartz cuvette (Starna Optiglass, Hainault, UK). Note that the sample was placed directly adjacent to the photomultiplier tube in order to minimize loss of light due to scattering. All measurements were performed in duplicate.

Rheology. Initial tests for gelation were carried out by simple vial inversion.[29] Typically, 2 mL of solution (with added GdL) was placed in a sample tube (22 mm diameter, 60 mm high). After standing for 24 h, the sample tube was inverted, with gelation being shown by a lack of flow. Dynamic rheological experiments were performed on an Anton Paar Physica MCR101 rheometer. In the oscillatory shear measurements, a sandblasted parallel top plate with a 50 mm diameter and 1.0 mm gap distance was used. Gels for rheological experiments were prepared on the bottom plate of the rheometer by loading a 2.0 mL solution of the gelator immediately after GdL addition. At this point, the sample is still a free-flowing liquid. Hence, sample uniformity and reproducibility are high. Evaporation of water from the hydrogel was minimized by covering the sides of the plate with low-viscosity mineral oil. The measurements of the shear modulus (storage modulus G' and loss modulus G'') with gelation were made as a function of time at a frequency of 1.59 Hz (10 rad/s) and at a constant strain of 0.5% for a period of 2 h, followed by an amplitude sweep from 0.01% to 100% under a frequency of 10 rad/s. The strain amplitude measurements were performed within the linear viscoelastic region, where the storage modulus (G') and loss modulus (G'') are independent of the strain amplitude. All the experiments were conducted at 25 °C.

Fluorescence Spectroscopy. Fluorescence spectra were obtained on a PerkinElmer luminescence spectrometer LS55. Thioflavin T (ThT) (0.1 mL of a 0.5 μM solution in water) was added to a solution of the dipeptide derivative (2 mL of a 0.5 wt % solution at pH 10.7). This was added to the required amount of GdL and immediately loaded into a 1.0 cm path-length cuvette. Emission spectra were collected every 5 min in the first 3 h and then every 30 min for 21 h. The emission spectra were recorded between 460 and 600 nm by excitation at 455 nm. The maxima emission peaks were at 485 nm. The slit width for emission and excitation was 2.5 nm, and the scan rate was 100 nm/min.

FTIR. IR spectra were collected on Bruker FTIR at 2 cm^{-1} resolution by averaging over 64 scans. All hydrogels were prepared in D_2O using NaOD (0.1 M) to change the pD. The hydrogels were loaded onto a CaF_2 windows, and another CaF_2 window was placed on the top of the gels. Each spectrum was background subtracted. For all spectra, the region between 1500 and 1800 cm^{-1} was shown in the figures after the peak of 1645 cm^{-1} was normalized to unity. The peak intensities were obtained by fitting the spectra with PEAKFIT software using Gaussian functions. The fitting coefficients were all above 0.98.

Transmission Electron Microscopy. Samples for examination by TEM were prepared *in situ* on Formvar/carbon film coated 400-mesh copper grids (Agar scientific). The required amount of GdL was added to a solution of dipeptide derivative (0.5 wt % solution at pH 10.7), and immediately grids were placed inverted onto 10 μL droplets of the gelation solution. In a humid chamber at room temperature, material was allowed to adsorb onto the grids and removed at specific time points followed by 5 min of drying and two 1 min negative stains using 2% w/v uranyl acetate. Grids to represent time zero were prepared by allowing adsorption of material for 1 min followed by negative staining as described. Negatively stained grids were allowed to dry and examined on a Hitachi-7100 TEM operated at 100 kV. Images were acquired digitally using an axially mounted (2000 × 2000 pixels) Gatan Ultrascan 1000 CCD camera (Gatan, Oxford, UK).

X-ray Fiber Diffraction. Fiber diffraction samples were prepared *in situ* from dipeptide solutions with varying amounts of GdL for final pH targeting. By placing a 10 μL droplet between two wax filled capillary tubes and allowing the solution to gelate and air-dry, a partially aligned fiber sample was formed (as previously described[30]). The fiber sample was mounted on a goniometer head, and X-ray diffraction data were collected using a Rigaku Cu Kα (wavelength 1.5419 Å) rotating anode and RAxis IV++ detector with exposure times of 10−20 min and specimen-to-detector distances of 200 mm. The diffraction images were examined using CLEARER,[31] and the diffraction signal positions were determined using a module within CLEARER.

(29) Raghavan, S. R.; Cipriano, B. H. In *Molecular Gels: Materials with Self-Assembled Fibrillar Networks*; Weiss, R. G., Terech, P., Eds.; Springer: New York, 2005.

(30) Makin, O. S.; Serpell, L. C. In *Amyloid Proteins: Methods and Protocols*; Sigurdsson, E. M., Ed.; Humana Press: Totowa, 2005; pp 67−80.

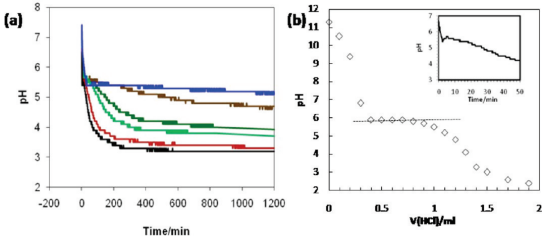
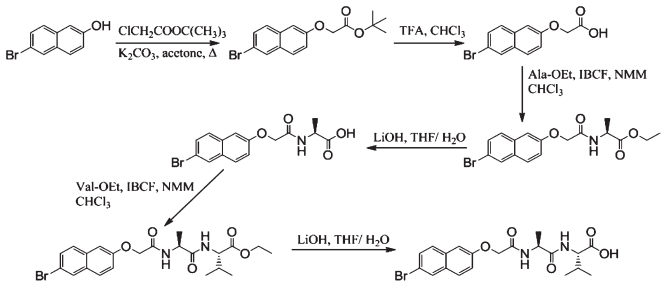


Figure 1. (a) Change in pH with time for aqueous solutions of dipeptide derivative (0.5 wt %) added to different amounts of GdL. In all cases, the initial pH was 10.7. Added GdL: (blue) 1.82 mg/mL, final pH = 5.1; (brown) 2.94 mg/mL, final pH = 4.7; (dark green) 4.46 mg/mL, final pH = 4.0; (light green) 5.96 mg/mL, final pH = 3.7; (red) 9.72 mg/mL, final pH = 3.3; (black) 14.42 mg/mL, final pH = 3.1. The final pH is quoted after 20 h. (b) Titration curve for the dipeptide solution with HCl (0.1 M) and pH changes curve in the first 50 min (inset) on adding dipeptide solution to GdL (14.42 mg/mL).

Scheme 1



Results and Discussion

The dipeptide derivative was synthesized from the commercially available 6-bromo-2-naphthol as shown in Scheme 1. Using this methodology, the naphthalene−dipeptide was recovered in high yield and purity (>98%, as evidenced by both NMR and elemental analysis).

This peptide derivative is related to the naphthalene−dipeptides reported by Yang et al.,[17] although here the naphthalene ring is also substituted with a bromine atom. This provides extra hydrophobicity compared to the parent naphthalene−dipeptide, which is a poor gelator (data not shown). An aqueous solution of the dipeptide derivative at a concentration of 0.5 wt % was prepared by addition of an equimolar amount of sodium hydroxide to a suspension of the dipeptide conjugate in deionized water. The final pH after peptide dissolution was found to be 10.7. The dipeptide derivative was found to be stable in aqueous solution at high pH for extended periods of time. To induce assembly, the solution was added to a specific amount of GdL. The kinetics of the pH change and the final pH of the solution were found to correlate with the amount of GdL used. Figure 1a shows that the pH drops rapidly initially, before reaching a pH of ∼5.4. Here, buffering occurs as the pK_a of the peptide derivative is reached. This value for the pK_a is higher than might be expected for the *C*-terminus of a peptide. However, it has recently been reported that the pK_a of the *C*-terminus of Fmoc-diphenylalanine is significantly higher than expected.[26] In line with this report, we carried out a "titration" of the dipeptide derivative with 0.1 M HCl solution. This confirmed that the pK_a of the dipeptide derivative was indeed higher than expected at around 5.9 (Figure 1b). Only a single pK_a was observed. With GdL, buffering occurs at a slightly lower pH of 5.4. It is also common to see a slight increase in pH after the initial buffering (the inset picture of Figure 1b). We ascribe this to the pK_a of the initial kinetic aggregates being higher than that of the dipeptide. Hence, on aggregation, if a relatively rapid change of the aggregates from carboxylate terminated to carboxylic acid terminated is quicker than the hydrolysis of GdL can provide further protons, an apparent pH increase will occur. The pH then continues to drop over time before reaching equilibrium.

For systems where the pH is below 5.0, self-supporting hydrogels were formed (as determined by simple vial inversion). The secondary structure of peptides (including dipeptides on gelation) has been probed previously by IR.[7,9,26,32] We measured the IR of gels prepared at a range of concentrations of GdL. The data in Figure 2a show that at high pH weak peaks at 1628 and 1679 cm^{-1} (normally ascribed to the presence of antiparallel β-sheets[32]) are observed. A stronger peak at 1645 cm^{-1}, which could be indicative of a random coil structure,[32] was observed. An

(31) Makin, O. S.; Sikorski, P.; Serpell, L. C. *J. Appl. Crystallogr.* **2007**, *40*, 966–972.

(32) Pelton, J. T.; McLean, L. R. *Anal. Biochem.* **2000**, *277*, 167–176.

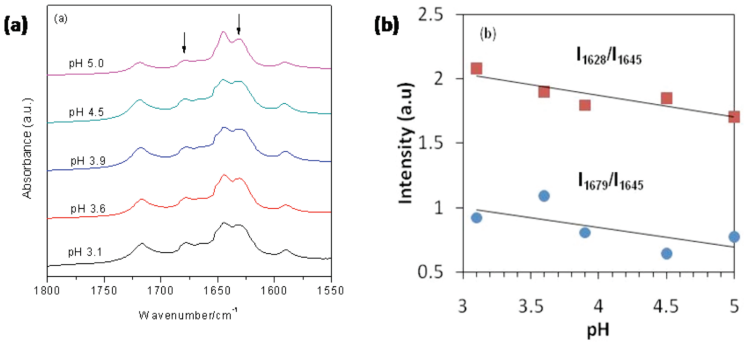


Figure 2. (a) IR spectra collected after 20 h from hydrogels prepared at a number of different pH. (b) Ratio of intensity of the peaks at 1629 and 1679 cm^{-1} relative to that at 1645 cm^{-1} increase as the pH decreases. The lines have been added as a guide to the eye.

alternative explanation for this peak is from the stretching of −C═C− in the naphthalene ring. At lower pH, the signals ascribed to the presence of β-sheet become stronger as compared to that at 1645 cm^{-1} (Figure 2b). As expected considering the pK_a of the dipeptide derivative, there is a peak at 1718 cm^{-1} present across the pH range measured, ascribed to the terminal COOH of the dipeptide. Recent data for Fmoc-diphenylalanine showed only the presence of two peaks, one at 1630 cm^{-1} and a second at 1685 cm^{-1}, both of similar intensity.[25,26] No peak at 1718 cm^{-1} was observed, implying that the terminal carboxylic acid is not protonated. These data were ascribed to the presence of antiparallel β-sheets. However, it is clear that the data reported for Fmoc-diphenylalanine, while similar to that reported here, are subtly different. This may indicate a slightly different packing arrangement of the dipeptide derivatives on assembly.

To probe further the interactions between the dipeptide molecules as the pH decreases, absorbance, CD, and LD spectra were recorded as a function time (which means as a function of pH in these experiments) following the addition of GdL to the solution to give a final GdL concentration of 14.42 mg/mL. The absorbance spectrum recorded for the dipeptide derivative at high pH has a very small signal above 300 nm, a medium strength signal at from 250 to 290 nm, and an intense band from 210 to 240 nm. Fine structure can be observed between 260 and 290 nm on careful inspection (Figure 3b). This structure is similar to that seen for naphthalene and substituted naphthalene compounds.[33] Thus, we can ascribe the 260−280 nm region to a naphthalene $\pi-\pi^*$ short axis polarized transition.[34] Similarly, much of the 230 nm intensity can be ascribed to an intense long axis polarized $\pi-\pi^*$ transition of the naphthalene. The peptide transitions are much weaker than those of naphthalene as shown by the decrease in absorbance between 200 and 220 nm, which is where the peptide bands have increasing intensity. Previously reported data for naphthalene−dipeptide conjugates showed a Cotton effect at 240 nm and a peak at 288 nm, ascribed to the $\pi-\pi^*$ transition of the naphthalene.[17]

Figure 3a shows the CD spectra of the naphthalene−peptide conjugates as a function of time (and hence pH). Naphthalene by itself is achiral and has no CD signal. In accord with this, the high-pH spectra show little or no CD signal when the conjugate is free in solution. However, as the pH drops, intense CD signals become increasingly apparent, indicating that the molecules are assembling into a chiral structure. The final CD magnitude (occurring at 60 min where the pH is measured to be 4.0 from Figure 1a, with no change thereafter) is high, indicating a highly asymmetric structure. The 260−280 nm short axis polarized naphthalene region gives a broad negative CD band with structure analogous to that observed in the absorbance spectra (Figure 3b, inset). At lower wavelength, the appearance of the spectrum is that of an overlay of exciton bands from 250 nm down. An exciton signal in a CD spectrum results from the coupling of degenerate transitions on nearby chromophores. The characteristic features of an isolated exciton band is that there are neighboring positive and negative bands with a fairly sharp transition between them whose zero point is at the maximum of the absorbance signal. When exciton bands overlay due to multiple transitions the picture becomes more complicated. In this case, the first exciton band is apparent as a large negative band at 236 nm and a positive component (apparent as a less negative minimum in Figure 3a) at 218 nm. Thus, the 230 nm band observed in the absorbance spectrum has split into a high-energy component and a low-energy component as a result of a chiral arrangement of the naphthalene chromophores.[35] If the chirality arises by a structure where the naphthalene chromophores are stacked one on top of the other, the sign pattern of the exciton CD spectrum can be interpreted following the line of argument in ref 35. The sign pattern observed here means that the long axes of two naphthalene chromophores are oriented[35] so the angle between them is between 0° and 90°, and they form a left-handed helix. As an aside, we note that the CD spectrum remains the same even if the sample is rotated through 90° (Figure 3d), thus verifying that the above discussion does relate to the true CD not a spectrum dominated by LD artifacts.

Recent data for a hydrogel prepared from Fmoc-diphenylalanine showed the presence of a strong negative CD signal at 218 nm, ascribed to the formation of a β-sheet structure.[25] No evidence was observed here for such a negative peak; however, this may simply be because the naphthalene chromophores

(33) Perkampus, H. *UV-Vis Atlas of Organic Compounds*, 2nd ed.; VCH Weinheim: New York, 1992.
(34) Rajendra, J.; Baxendale, M.; Rap, L. G. D.; Rodger, A. *J. Am. Chem. Soc.* **2004**, *126*, 11182–11188.
(35) Rodger, A.; Norden, B. *Circular Dichroism and Linear Dichroism*, 1st ed.; Oxford University Press: New York, 1997.

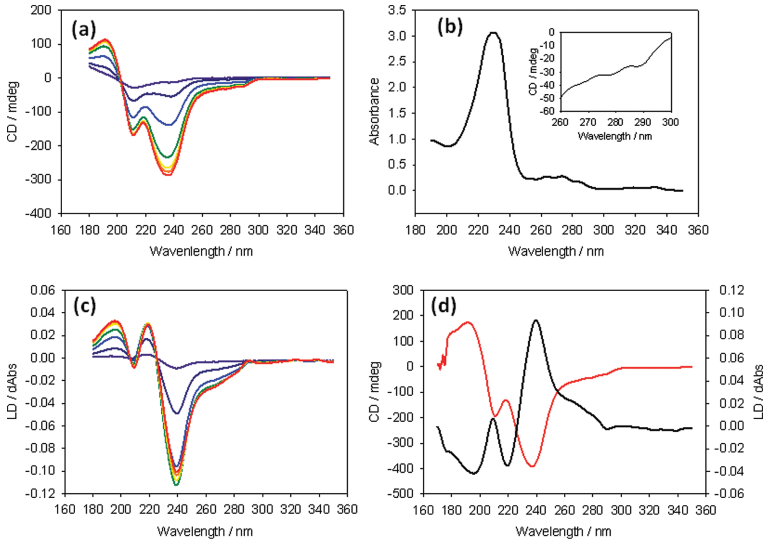


Figure 3. (a) Evolution of CD with time on addition of a solution of dipeptide derivative (0.5 wt %) to GdL (14.42 mg/mL). Data are shown for 0 min and then subsequently every 10 min. Signals increase with time. (b) UV−vis spectrum for dipeptide derivative at pH 10.7. Inset is an expansion of the CD after 1 h showing fine structure between 290 and 270 nm. (c) Evolution of LD with time on addition of a solution of dipeptide (0.5 wt %) to GdL (14.42 mg/mL). Data are shown for 0 min and then subsequently every 10 min. Signals increase with time until 50 min, when the magnitude decreases once again. (d) CD (red) and LD (black) spectra for sample formed 60 min after GdL addition with cell rotated through 90° (pH = 4.0 from Figure 1a).

dominated the observed spectrum due to their high extinction coefficient and consequent effective $\pi-\pi$ coupling.

The CD signal of $\pi-\pi$ systems is dominated by nearest-neighbor couplings. In order to try to determine how the naphthalene chromophores were oriented in the extended structure we assumed was present in the hydrogel, we measured the LD as a function of time (and hence pH) since it reports on the macroscopic anisotropy. LD is the difference in absorbance of light polarized parallel and perpendicular to an orientation axis. Samples were prepared in the cuvettes as before, with any alignment arising simply from structure formation in the cell. Here, the spectrum again changed over the first 40 min, with the magnitude of the signals increasing, before beginning to decrease again at 50 and 60 min (Figure 3c). The magnitude of the LD signals depended on the sample, with variation being observed between nominally identical samples (unlike the CD spectra which were of very similar intensity for different samples). As expected, the LD spectrum is inverted on rotating the sample through 90° (Figure 3d). From the CD band signs we know that the 236 nm signal results from the out-of-phase coupling (thus perpendicular to the line between the long axes of the napthalenes) of the transition moments[35] and the 218 nm signal from in-phase coupling. If the naphthalenes are stacked vertically like steps on a ladder, then both transitions would be expected to have a negative LD. This is in contrast to what is observed here. This means that the naphthalene long axes are tilted by more than 35° from the perpendicular. (LD = $3\cos^2\alpha - 1$, where α is the angle between the orientation axis and the transition moment. As LD = 0 for α = 54.7°, we deduce that the tilt must ensure the naphthalene's are tilted so they lie less than 54.7° from the orientation axis of the sample.)

To follow the assembly and gelation process, thioflavin T (ThT) was used as a fluorescent probe. ThT is widely used to probe the formation of β-sheet structures, especially in amyloid-forming peptides.[36,37] Recent results suggest that ThT acts as a molecular rotor.[38] Molecular rotors are viscosity-sensitive molecules, with a well-defined relationship between viscosity and fluorescence quantum yield.[39] Hence, an increase in local viscosity results in a significant increase in observed fluorescence. Liang et al. have demonstrated that Congo Red, a stain for amyloidogenic peptides, can also be used to stain fibers prepared from naphthalene−di- or tripeptides.[40] Since the IR data above demonstrate the presence of β-sheet structures, we rationalized that ThT could be used to probe the local structure as gelation occurred. ThT was therefore directly incorporated in the stock solution of the dipeptide derivative at pH 10.7. The lack of change in fluorescence at early time points indicates that assembly only begins when the pH reaches the pK_a of the dipeptide derivative (Figure 4): no increase in fluorescence is observed until a pH of 5.4

(36) Hawe, A.; Sutter, M.; Jiskoot, W. *Pharm. Res.* **2008**, *25*, 1487–1499.
(37) Lindgren, M.; Sorgjerd, K.; Hammarstrom, P. *Biophys. J.* **2005**, *88*, 4200–4212.
(38) Stsiapura, V. I.; Maskevich, A. A.; Kuzmitsky, V. A.; Uversky, V. N.; Kuznetsova, I. M.; Turoverov, K. K. *J. Phys. Chem. B* **2008**, *112*, 15893–15902.
(39) Haidekker, M. A.; Theodorakis, E. A. *Org. Biomol. Chem.* **2007**, *5*, 1669–1678.
(40) Liang, G. L.; Xu, K. M.; Li, L. H.; Wang, L.; Kuang, Y.; Yang, Z. M.; Xu, B. *Chem. Commun.* **2007**, 4096–4098.

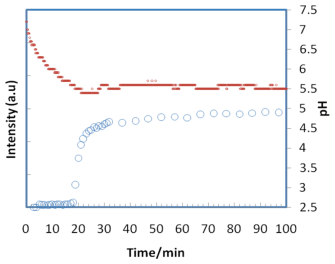


Figure 4. Change in ThT fluorescence at 485 nm (λ_{ex} = 455 nm) on addition of a solution of dipeptide derivative (0.5 wt %) to GdL (2.94 mg/mL) (blue data). Overlaid is the change of pH with time (red data from Figure 1).

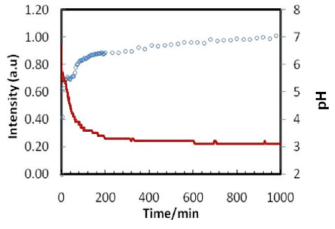


Figure 5. Normalized change in ThT fluorescence at 485 nm (λ_{ex} = 455 nm) on addition of solutions of dipeptide derivative to GdL (blue data). Overlaid is the change in pH with time (red data). The initial pH was 10.7. Amount of GdL added 14.42 mg/mL; final pH = 3.1.

is reached, following which a rapid increase in fluorescence is observed.

The change in ThT fluorescence over a longer time resulting from the addition of GdL to a solution of the dipeptide derivative is shown in Figure 5. From these data, it can be seen that there is a rise to a plateau (that observed in Figure 4), followed by a second rise to a plateau. This behavior is exhibited at a number of GdL concentrations (see Supporting Information).

Interestingly, the initial increase in fluorescence to the first plateau occurs while the systems are still liquid (see below for rheology data in Figure 9). This is in agreement with data obtained previously for Fmoc−leucine−glycine,[10] where fibers were imaged at early times before gelation occurred. The data obtained here indicate that assembly occurs by a two-stage process. Further analysis demonstrates that while the time at which the first plateau is reached is determined by the amount of GdL added (and hence the kinetics of pH change), the absolute pH at which this transition occurs is extremely similar in each case at pH 5.2 (see Supporting Information). This indicates that this transition is dependent on the pH of the solution, rather than the time taken to reach this pH. The end of the plateau occurs at different pH values depending on the amount of GdL added but occurs more quickly at higher quantities of GdL.

Using this information, the system was further probed by transmission electron microscopy (TEM). A solution containing 14.42 mg/mL of GdL was prepared. Aliquots were removed and allowed to gel for defined times. At these times, the samples were analyzed by TEM. The time course in Figure 6 shows that immediately after the addition of GdL ill-defined structures are formed. After 40 min (pH 4.2 from Figure 1a), some fibers with a minimum width of 12.9 nm (SD ± 0.97 nm, n = 10) were observed. After 80 min (pH 3.8), a large number of fibers were imaged. There is clear evidence for lateral association between fibers. Thicker structures are formed via the association of thinner fibers with a width of 12.0 nm (SD ± 0.50, n = 10) (Figure 7). The associated fibers have a higher persistence length than the thinner fibers. The micrographs taken at later times appear very similar showing a network of fibers, the majority of which are thicker than the original fibers formed at earlier time points. In contrast to results previously reported for naphthalene−dipeptides,[17] no helical structures were observed.

TEM was also used to probe the structures formed after 24 h in the systems with different amounts of added GdL. Here, the role of the final pH of the sample was examined. As can be seen from Figure 8, at a final pH of 5.0, the fibers formed are narrow with a minimum width of 10.5 nm (SD ± 1.53, n = 10); these fibers are comparable to the thinnest fibers observed at early time points (Figure 6). These fibers are likely to represent a basic protofilament. At pH 4.5 and below, similar structures were imaged but some of them with larger widths, in one instance measuring 25.4 nm (SD ± 1.83, n = 10). These also look very similar to the final structures prepared during the time course experiment (Figure 6) and are indicative of laterally associated protofilaments.

From the collective data shown above, it is clear that the self-assembly of the dipeptide derivative occurs via $\pi-\pi$ stacking and hydrogen bonding as would be expected. β-Sheet formation occurs, with the apparent degree of β-sheet increasing at lower pH. We have shown how GdL can be used to control the pH of the solution in a uniform fashion. The concentration of GdL added determines the final pH, and time determines how far along the route to the final pH that the solution is. The kinetics of assembly are determined by the rate of pH change in these experiments. In addition, the nature of the final hydrogel formed is dependent on the final pH of the solution, which we can also control by adding the appropriate amount of GdL to the starting mixtures. From these data, it is apparent that the kinetics of hydrogel formation and final material properties of the hydrogels would be expected to correlate with the amount of GdL added since this determines pH. Control of the material properties is key for final applications. The mechanical properties of the hydrogels prepared using different amounts of GdL were therefore probed by rheology. As reported previously for Fmoc−dipeptides and amino acids with GdL,[10] the development of the gel structure with time was probed *in situ*. Figure 9 shows the time evolution of storage modulus (G') and loss modulus (G'') with time for the gelation process of the dipeptide derivative using different amounts of GdL. In all cases, the gel strength (as measured by the storage modulus, G') increases with time before coming to a plateau value. Initially, the storage modulus and the loss modulus are similar in value as expected for a liquid. With time, G' dominates over G'' as expected for a true hydrogel. It should be noted that the gelation point (i.e., where $G' = G''$) occurred at a very early stage (less than 30 min). The kinetics of the development of the gel network depends on the amount of GdL, correlating well with the kinetics of the pH change. However, the final plateau values for both G' and G'' are extremely similar. The exception is for the lowest concentration of GdL, where a slightly lower value was obtained. However, the values for the modulii after 48 h were very similar to those shown for the higher GdL concentrations (data not shown). These data indicate

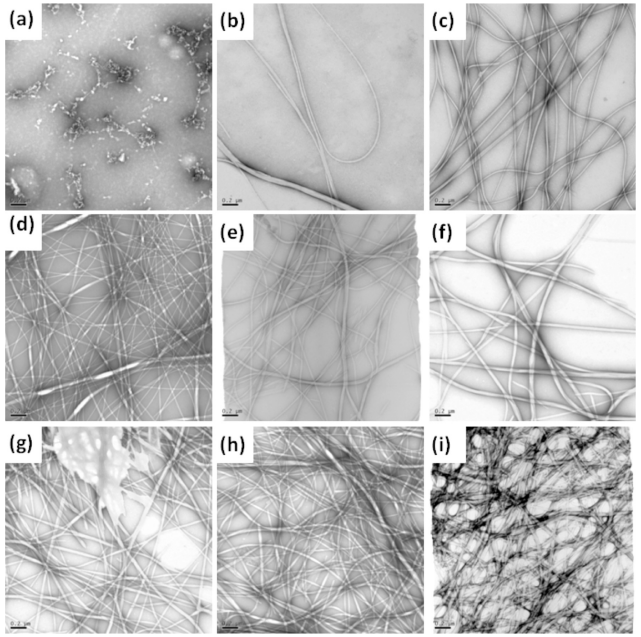


Figure 6. TEM of evolution of structures with time for dipeptide in the presence of GdL (14.42 mg/mL): (a) immediately after GdL addition; (b) 40 min, (c) 80 min, (d) 120 min, (e) 160 min, (f) 200 min, (g) 240 min, (h) 280 min, and (i) 400 min after GdL addition. In all cases, the scale bar represents 200 nm.

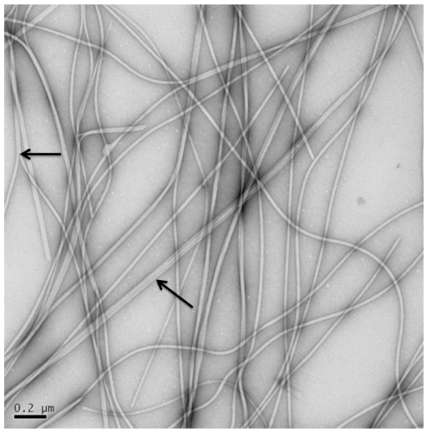


Figure 7. TEM of fibers formed 80 min after GdL addition. Associated fibers are highlighted with a black arrow. The scale bar represents 200 nm.

that the three-dimensional network structures that contribute to the mechanical properties are very similar in all cases after 24 h, despite the different kinetic pathway by which these networks are formed. We also note that the final value for the storage modulus (G') of ~57000 Pa is significantly greater than that previously reported for a naphthalene−dipeptide (~8000 Pa).[17] We ascribe this to either the use of GdL to change the pH uniformly across the reaction mixture, thus achieving a uniform hydrogel, since a similar effect was noted for Fmoc−dipeptide systems,[10] or the higher hydrophobicity of the dipeptide derivative used here.

Among all material viscoelasic functions, complex viscosity (η^*) is considered as another sensitive parameter to structural changes during gelation.[41] Figure 10 shows the time dependence of η^* overlaid with the intensities of emission peaks of ThT during the gelation process for systems using different concentrations of GdL. This data demonstrate that the evolution of structure as demonstrated by rheological measurements correlates well with the fluorescence data. In all cases, a two-stage process occurs, with the kinetic profile being dependent on the amount of GdL added. Similar to the behavior of intensities of ThT emission peaks with time, η^* increased with the gelation. They also show two-stage evolutions.

Despite similarities in G' and G'' between different systems, the impact of the kinetics of gel formation can be observed in the

(41) Weng, L. H.; Chen, X. M.; Chen, W. L. *Biomacromolecules* **2007**, *8*, 1109–1115.

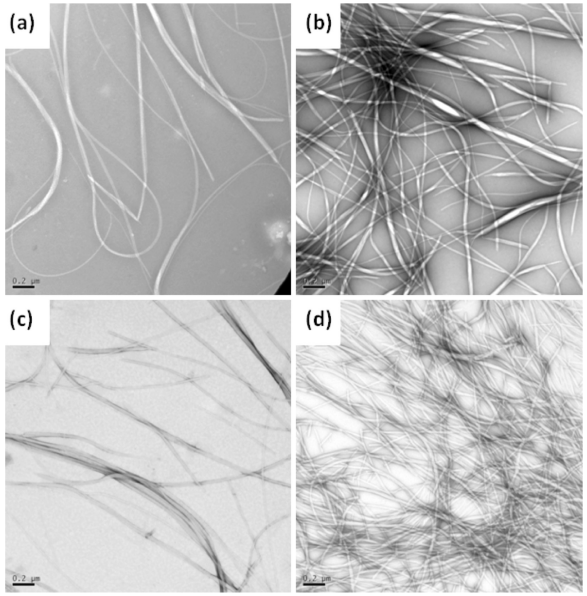


Figure 8. TEM of fibers formed after 24 h in dipeptide hydrogels (0.5 wt %) with different amounts of added GdL: (a) 1.82 mg/mL, final pH = 5.0; (b) 2.94 mg/mL, final pH = 4.5; (c) 5.96 mg/mL, final pH = 3.6; (d) 14.42 mg/mL, final pH = 3.1. The scale bar in each case represents 200 nm.

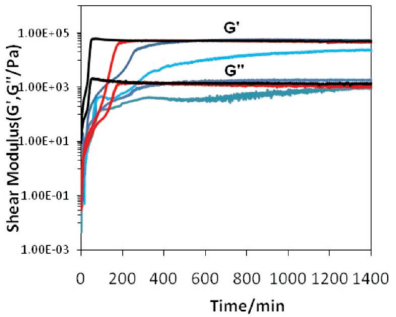


Figure 9. Evolution of G' and G'' with time for solutions of the dipeptide (0.5 wt %) with different amounts of GdL. Concentrations of GdL used: (black) 14.42 mg/mL, (red) 5.96 mg/mL, (dark blue) 4.46 mg/mL, and (light blue) 2.94 mg/mL.

amplitude sweeps for the hydrogels shown in Figure 11. Here, the hydrogel prepared using 14.42 mg/mL GdL (i.e., final pH 3.1) was found to be the most stable under external force, with the transition from "gel-like" to "liquid-like" (as demonstrated by the crossover point where G' is no longer greater than G'') is around 5%. At which, there is a rapid decline of G', indicating the breakdown of hydrogel structures. This is similar to that reported elsewhere for gels prepared using di(*p*-toluoyl)-L-cystine.[42] The gels formed with less GdL at a higher final pH (pH = 4.0 and 4.5) are less tolerant to strain, with the turning points occurring at ~1%. However, the crossover point from "gel-like" to "liquid-like" occurs at much higher strain in these systems at approximately 8−10%. This failure behavior implies that the fibrils forming the matrix of the hydrogels formed at lower pH are very rigid and so unable to withstand large deformations. This is similar to that reported for other related systems.[10] The data for the gels prepared at lower GdL concentrations and hence higher final shows that the gels still have elastic gel-like structure until a strain of 8%, although the three-dimensional networks have already been broken under the strain of 1%.

To gain further insight linking the rheology to the local packing of the peptides, we prepared fibers from gels prepared using different amounts of GdL and hence different final pH. Comparison of the X-ray fiber diffraction patterns collected from "*in situ*" prepared fiber alignments indicates that all gels formed at all final pH values have the same underlying molecular architecture. The patterns obtained appear to exhibit reflections similar to those observed for the cross-β patterns observed from amyloid, showing a meridional reflection at 4.5 Å (Figure 12) corresponding to β-strand spacing along the fiber axis. The reflections observed are similar for the samples at different pH (also see Figure S2, Supporting Information). The patterns differ only in the extent

(42) Menger, F. M.; Caran, K. L. *J. Am. Chem. Soc.* **2000**, *122*, 11679–11691.

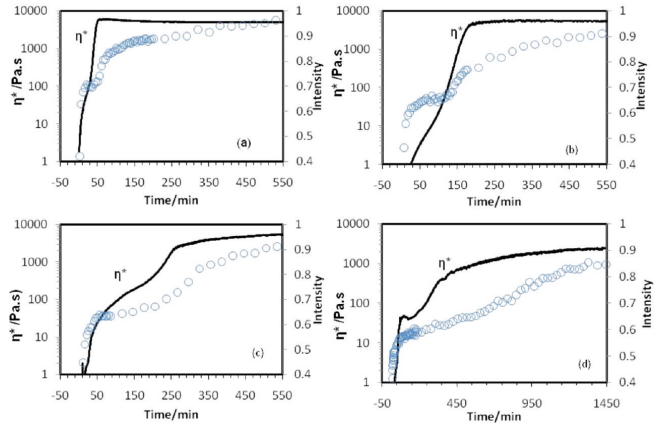


Figure 10. Relationship between complex viscosity η^* and fluoresence intensities for hydrogels prepared using different concentrations of GdL: (a) 14.42, (b) 5.96, (c) 4.46, and (d) 2.94 mg/mL.

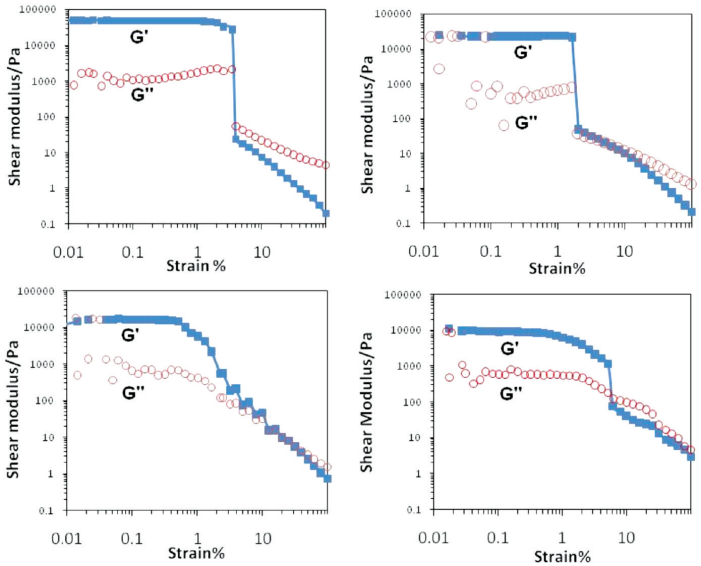


Figure 11. Strain sweeps for hydrogels prepared using different concentrations of GdL: (a) 14.42, (b) 5.96, (c) 4.46, and (d) 2.94 mg/mL. All data were collected at a frequency of 10 rad/s.

of alignment (shown by the degree of distinction between meridion and equator). In general, as pH is lowered orientation diminishes, as indicated by the major meridional reflections changing from well-defined arcs (pH 5.0) to rings (pH 3.1). These two observations are consistent with the behavior of the gels under strain as seen in the rheology measurements and the lateral association observed in the TEM. The data can describe a model whereby the molecular arrangement in the fibers is adopted rapidly at pH 5.0 (< 40 min), but the gel strength is low enough that these fibers can be aligned. As the pH continues to fall, the

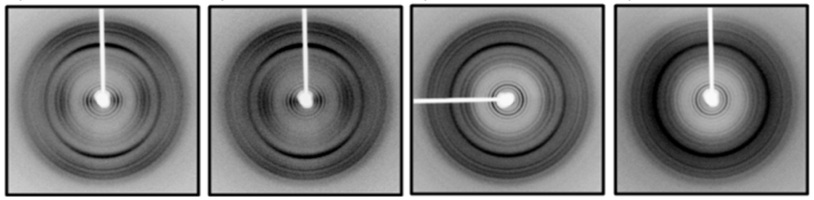


Figure 12. X-ray fiber diffraction patterns collected from "*in situ*" prepared fiber alignments from gels at a final pH from left to right of 5.0, 4.6, 3.6, and 3.1. All fiber axes are vertical to the diffraction patterns. The major meridional reflection is found to be at 4.5 Å and first major equatorial reflection found at 16 Å followed by a second grouping of reflections starting at 7.3 Å.

underlying molecular arrangement is unchanged but gel strength increases due to the increased lateral association of fibers.

Synchrotron fiber diffraction data showed the presence of two low-resolution equatorial reflections at 38 and 27 Å (data not shown). All the diffraction signal positions were examined using CLEARER, and a possible unit cell was determined.[31] Diffraction signals were indexed to the most likely unit cell (see Table S1, Supporting Information). The cell dimensions obtained were a = 37.90, b = 27.28, and c = 4.51 Å. The meridional 4.5 Å reflection is likely to arise from strand spacing along the fiber axis, as observed in amyloidogenic cross-β systems, and a similar strand spacing has been assigned in other dipeptide systems.[25] The equatorial reflections in the patterns arise from the unit cell dimensions perpendicular to the fiber axis and were used for unit cell determination.

We find that the unit cell may allow for more than one potential packing arrangement whereby naphthalene stacking is parallel to either the a or b axis. Many of the diffraction signals arise from higher order reflections from the unit cell itself (Table S1, Supporting Information), but the signals at 7.3 and 5.4 Å can be indexed as arising from a repetitive spacing within the unit cell, which we believe to result from naphthalene stacking, backed up by the CD data discussed above (Figure 3). We find it interesting to note that Smith et al. describe for the Fmoc−diphenylalanine system a β-strand spacing of 4.3 Å and a distance between pairs of aromatic fluorenyl groups of 7.6 Å,[25] very similar to what we observe here.

Conclusions

Combining all of these data, it is clear that the kinetics of self-assembly of the dipeptide derivative and hence hydrogelation is determined by the kinetics of pH adjustment. Using GdL, it is possible to adjust the kinetics of the pH drop, and hence we can probe the assembly process using a number of techniques. The importance of the kinetics of assembly has been described elsewhere for a temperature triggered assembly of aroyl L-cystine derivatives.[42] Similarly, the importance of history has been discussed for organogelators.[43] However, for the system described here, the assembly process requires a change in ionization state of the gelator. From the data, assembly begins as charge is removed from the C-terminus of the dipeptide derivative. The pK_a is higher than might be expected for the C-terminus of a dipeptide. However, this is in agreement with a recent report for Fmoc−diphenylalanine. IR shows that β-sheet assemblies are formed as the pH drops. The CD and LD data demonstrate that the assembly of the dipeptide derivative leads to a chiral organization of the naphthalene rings. This assembly occurs at relatively high pH (i.e., early on in the self-assembly process), with no increase in signal intensity after a pH of ~4. From the fluorescence data with ThT, it is clear that the assembly occurs via a two-stage process—the first stage occurring at higher pH (~5.2). This correlates well with the pK_a of the dipeptide derivative, indicating that the first stage of the assembly is associated with the removal of charge on the C-terminus of the dipeptide. This removal of charge allows lateral association of the dipeptides and hence β-sheet formation. Following this, a second stage of the assembly occurs, the kinetics of which are determined by the concentration of GdL added. This assembly leads to a matrix that has very similar rheological properties (as measured by G' and G''), indicating that this is not affected by the final pH. This is unsurprising—the only pH-sensitive group is the C-terminus, which will be fully protonated during this assembly. It may be that the kinetics of the assembly is driven by the kinetics of the primary assembly process, rather than the absolute pH. Indeed, it has been shown elsewhere for hydrogels prepared from a three-stranded β-sheet forming peptide that more rigid hydrogels are formed at higher temperatures (the trigger in this case).[44] Hence, faster kinetics of assembly leads to a more rigid gel. This was ascribed to either morphological differences in the fibrillar network (e.g., more physical cross-links) or alternatively intermolecular interactions that are dependent on the temperature of assembly. Similar effects may be occurring here. For the dipeptide studied here, the final mechanical properties of the hydrogels are very similar irrespective of the amount of GdL added; rather, the time to achieving the final gel is determined by the GdL concentration. The kinetics of the assembly has a clear effect on the ability of the networks to withstand strain, implying that it is the cross-links between fibers that are different, rather than the fibers themselves. This is confirmed from the X-ray fiber diffraction data. This data gives great insight into the assembly mechanism, demonstrating how assembly leads to hydrogelation.

Acknowledgment. We thank the EPSRC (EP/G012741/1) for funding.

Supporting Information Available: Further ThT fluorescence data at different GdL concentrations, one-dimensional graphical representation of the equatorial axes from X-ray fiber diffraction patterns, and comparison of the experimentally observed diffraction signals with the predicted signals from the determined unit cell. This material is available free of charge via the Internet at http://pubs.acs.org.

(43) Huang, X.; Raghavan, S. R.; Terech, P.; Weiss, R. G. *J. Am. Chem. Soc.* **2006**, *128*, 15341–15352.

(44) Rughani, R. V.; Salick, D. A.; Lamm, M. S.; Yucel, T.; Pochan, D. J.; Schneider, J. P. *Biomacromolecules* **2009**, *10*, 1295–1304.

From natural to designer self-assembling biopolymers, the structural characterisation of fibrous proteins & peptides using fibre diffraction†

Kyle Morris[a] **and Louise Serpell***[b]

Received 19th April 2010
DOI: 10.1039/b919453n

Protein and peptide self-assembly is integral to a growing number of misfolding diseases that involve the ordered assembly and deposition of misfolded proteins as amyloid fibrils. However, it is becoming clear that these highly stable assemblies can be exploited as potential bionanomaterials. Recent work reveals that amyloid-like assembly is carefully controlled by a number of organisms to form functional amyloid. It is essential to gain an understanding of how sequence influences the assembly mechanism, as well as the final structure, to enable the control that is required to fully exploit these materials. In this *tutorial review*, we will discuss the different types of fibrous protein assembly from structural and cytoskeletal proteins through to misfolded proteins and finally designed self-assembling peptides. In the second part, we will discuss the advances in structure determination, with particular reference to the use of X-ray fibre diffraction.

Introduction to peptide self-assembly

Fibrous structures are central to many biological functions and are increasingly being realised as useful nanomaterials.[1–4] The route to the production of such structures is most easily available through self-assembly. The self-association and assembly of proteinaceous material is a recurring theme in biological systems. Non-covalently linked assemblies can be formed from small molecules with molecular weights as low as 450 Da to large polypeptides. In recent years, the self-assembly of proteins to form β-sheet rich fibrillar structures known as amyloid fibrils has highlighted the ability of many short peptides and large folded proteins, to misfold and self-assemble to form very similar underlying structures.[5] Understanding the mechanism of assembly and the structure of these assemblies is of importance for understanding fundamental biological structures such as the cytoskeleton and for treating neurodegenerative diseases, through to designing a new range of biomaterials for future biological and technological applications.

Several strategies are available for the synthesis of a large macromolecular system[6] but a self-assembling system has a number of appealing features. The technical challenge of creating a relatively small monomer is much less than the final assembled structure. Where the assembly is stabilised by non-covalent forces then its polymerisation and depolymerisation are potentially controllable. Additionally because the self-assembly process must be spontaneous it represents a thermodynamically favourable transition, the structures formed occupy a minimum making them inherently stable.[6] Fig. 1 illustrates that a wide range of monomeric structures can all assemble into fibrillar macromolecules. However the challenge in making such a system lies in designing the self-assembly mechanism, where

[a] *School of Life Sciences, John Maynard-Smith building, University of Sussex, Brighton, UK. E-mail: km72@sussex.ac.uk; Tel: 012 7387 3365*
[b] *School of Life Sciences, John Maynard-Smith building, University of Sussex, Brighton, UK. E-mail: lcs21@sussex.ac.uk; Tel: 012 7387 7363*
† Part of the peptide- and protein-based materials themed issue.

Kyle Morris

Kyle Morris completed his BSc at the University of Sussex in Molecular Medicine. He is now working for his PhD thesis under the supervision of Louise Serpell. His research involves the structural and biophysical characterisation of a broad range of amyloid-like self-assembling systems. The focus of this work is X-ray fibre diffraction, development of this technique and the analysis required.

Louise Serpell

Louise Serpell is currently a Reader in Biochemistry at the University of Sussex. She specialises in the structural aspects of amyloidogenic peptides, utilising X-ray fibre diffraction and electron microscopy. She obtained a DPhil from the University of Oxford and then moved to the University of Toronto for a postdoctoral position. Louise returned to the UK to take up a position at the Laboratory of Molecular Biology in Cambridge and then held a Wellcome trust fellowship at the University of Cambridge.

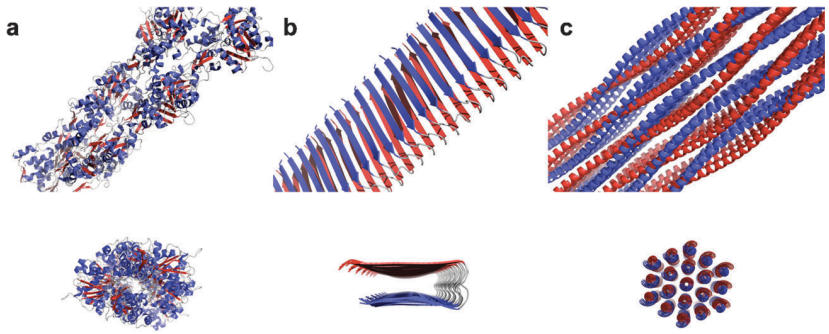


Fig. 1 Fibrillar assemblies can be formed from a diverse range of monomer subunits including the mixed secondary globular structure of two actin filaments[48] (a), cross-β laminated sheets of Aβ(1–40) amyloid[69] (b), and filaments formed by self-assembled α-helical coiled-coils,[36] known as SAFs (c). Illustrations are shown diagonal and perpendicular (top) and parallel to the fibril axis (bottom). Molecular graphics generated using Pymol.[70]

currently the principles of non-covalent forces are well understood in isolation but the application of these is an emerging science. Table 1 summarises the systems described here. To some extent all are biologically related or inspired, though some have sought to fully recreate their biological mimics and others have taken key features in isolation to create a new system.

Cellular fibrous proteins

Fibrous proteins are centrally important in biological systems, providing support for larger structures or making up the effective parts of protein motor systems. A self-assembled proteinaceous fibril is an ideal structure for a cellular support. Mechanical strength is derived mainly from a high density of lateral and longitudinal non-covalent bonds in the body of the fibril and, in the case of cytoskeletal elements, differences in the nature of these interactions at the termini afford the fibril the ability to dynamically de/repolymerise[7] permitting the reorganisation of the structures that it supports. Components of the cytoskeleton that provide structural support and motile forces to eukaryotic flagella and cilia include filamentous microtubules,[8] additionally actin filaments play a vital role in the formation of filopodia during neuronal growth.[9] Microtubules and actin filaments consist of self-associated monomers of tubulin[10] and actin[11] respectively. Tubulin and actin have guanidine and adenosine triphosphate hydrolytic activity that gives rise to conformational states favouring self-assembly. In this manner assembly can be controlled, and this is key because a number of diseases are related to the uncontrolled self-assembly of proteins. The mixed secondary structure of fibrillar actin is shown in Fig. 1a, these fibrils have a width of ~10 nm and a helical morphology as shown in the electron micrograph in Fig. 2a.

Table 1 A summary of a range of fibrillar proteinaceous systems and their constituent monomer

Fibrillar system	Monomer	Secondary structure composition
Actin	Actin	α-helix/β-sheet
Silk	$[G\text{-}A]_n$	β-sheet crystallites
	$[A]_n$	
Amyloid	Aβ	Cross-β
	α-synuclein	Cross-β
	Poly-Q	Cross-β
	IAPP	Cross-β
	PrP	Cross-β
	TTR	Cross-β
Amyloid-like	Sup35	Cross-β
	WALTZ peptides	Cross-β
De-novo designed Hydrogels	Amphiphilic β-strands	β-sheet
	FMOC-FF	β-sheet
	Naphthalene-XX	β-sheet
	hSAFs	α-helix

It has recently been shown that an amyloid-like structure is central to some functional fibrillar structures. A key example is *Escherichia coli* (*E. coli*) curli fibrils, involved in cellular adhesion. The assembly of this system is carefully controlled by the expression of two proteins, CsgA and CsgB. CsgA constitutes the fibril that *in vitro* can spontaneously self-assemble but *in vivo* requires the presence of CsgB to nucleate fibril growth.[12] This precise control over fibril formation is unique amongst amyloid systems. Systematic alanine scanning of CsgA reveals specific terminal residues critical for assembly. *In vitro* spontaneous self-assembly and *in vivo* nucleation by csgB are negated when these residues are changed, however mutants may still be seeded by pre-existing csgA fibrils, alluding to a distinct assembly mechanism.[13] Two key residues are solely responsible for *in vivo* aggregation and this demonstrates that mechanisms may be designed to control assembly in a very specific manner.

Synthetic silk

Although natural silk is typically a composite material[14] it has a proteinaceous core rich in β-sheet assemblies. It is understood that the exceptional mechanical strength, in some cases outperforming all man-made materials, of a silk fibre is derived from this core. The basic stabilising unit is the hydrogen bond. Comparatively weak compared to other interactions and certainly a fraction of the strength of a covalent bond, the intrinsic weakness of the hydrogen bond is overcome by using structural hierarchies. Molecular dynamics simulations have suggested that the hydrogen bonding networks of assembled β-sheets when stressed will only break in groups and it is this phenomenon that affords such assemblies their great strength.[15] The high tensile strength of spider silk indicates its potential for materials applications. Select proteins from the dragline silk of the spider *Araneus diadematus* are able to form assembled β-sheet architectures by way of polyalanine repeats.[4] Two engineered proteins, eADF3 and 4, based on the proteins comprising silk, self-assemble into spherical aggregates and under particular ionic and microfluidic mechanical conditions assemble into fibres.[16] The production of these synthetic silks not only represents a significant technological advancement in the uses of microfluidic systems and an addition to the materials available to the biomaterials scientist but a promising system for further understanding the self-assembly process. In these silks, the small side chains of the alanines allow tight packing of the intermolecularly hydrogen bonded β-strands and this appears to be an important factor contributing to stable fibre formation. Indeed a number of $[G\text{-}A]_n$ and $[A]_n$ silk inspired assembling systems have been reported (for review see ref. 17).

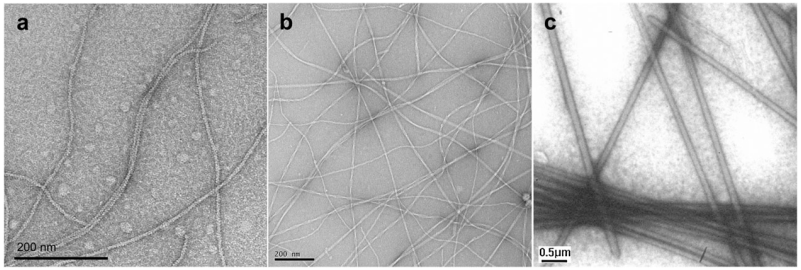


Fig. 2 Transmission electron micrographs of actin (a), mature Aβ-amyloid fibrils (b) and coiled-coil SAFs[36] (c).

Amyloid

The subject of a great deal of study over many years, a self-assembling protein has been placed as the central cause of Alzheimer's disease.[18] The spontaneous aggregation of proteins into higher order structures, known as amyloid, is in fact associated with a number of degenerative diseases. The amyloid fibrils share a β-sheet rich architecture known as cross-β structure (Fig. 1b) and have varied morphology with a width of 10–20 nm and an indeterminate length (Fig. 2b). Each disease is associated with a different assembling protein, for example Parkinson's disease is associated with the aggregation of α-synuclein, Huntington's disease with expanded polyglutamine caused by CAG repeats, Alzheimer's disease by an aggregation of amyloid-β peptide (Aβ), Type 2 diabetes with the accumulation of islet amyloid polypeptide and a number of transmissible spongiform encephalopathy prion diseases with the PrP protein.[19] Whilst these diseases may be characterised by the presence of their corresponding aggregated proteins, the precise correlation between aggregation and disease pathology is not entirely clear and still under investigation.[20]

Amyloid-like fibrils may be formed from synthetic or recombinant proteins *in vitro*. These systems range from full length or short fragments of amyloidogenic proteins, sequences identified as having amyloidogenic propensity by a bioinformatics approach or designed in a 'bottom up' manner based rationally on sequence physicochemical properties. A great number of amyloid fragments have been investigated with the aim to understand disease-related amyloids formed by full length proteins.[5,21,22] It is reasonable that in addition to this, any structure function relationships and molecular features that influence supramolecular characteristics established could be used for the rational design of biomaterials.

The amyloidogenic proteins associated with disease share very little obvious sequence similarities. Indeed it has been suggested that all proteins and peptides share the ability to assemble to form amyloid-like fibrils under the appropriate denaturing conditions.[19] In-depth investigation of the propensity of different sequences to aggregate has been performed by a number of groups, focussing initially on using overall physicochemical properties. Sequences have successfully been identified that aggregate[23–27] but do not necessarily have the same inherent order as amyloid fibrils. Subsequently, ordered amyloid assembly[7,28,29] has been examined by using sequence specificity and residue order. Recently short peptide motifs have been identified that have the ability to form highly ordered amyloid-like structures[29] and fibrous crystals where this work has highlighted the importance of associations that form stabilised "steric zippers"[22] referring to the close interdigitation of side chains.

The structural determination by MAS-NMR of a peptide in an amyloid fibril, a fragment of the amyloidogenic protein transthyretin (TTR), revealed that TTR amyloid peptide molecules adopt an extended β-strand conformation. The fibrillar peptides are positioned in amyloid fibrils with such great regularity, long range order and limited variation in dihedral angles that their degree of order is comparable to that only found in crystals.[30] This affords the efficient formation of

inter-backbone amide hydrogen bonding parallel to the fibre axis and thus strength. Another system found to form highly ordered crystalline fibres is that of a polyglutamine sequence (poly-Q). A molecular model of poly-Q in the compact crystalline state has been proposed, whereby the characteristic hydrogen bonding pattern is complemented by contributions from side chains.[31] The peptide $D_2Q_{15}K_2$, representative of the pathological expanded polyglutamine protein huntingtin, consists of a repetitive glutamine core (Qblock) flanked by two aspartates and two lysines. The insoluble crystals produce an X-ray fibre diffraction pattern characteristic of the cross-β architecture. Similarly, the model proposes β-sheets of $D_2Q_{15}K_2$ are stabilised by inter backbone amide hydrogen bonding parallel to the fibre axis but uniquely by additional hydrogen bonding from glutamine residues in which one glutamine acts as a donor while its neighbour acts as an acceptor. Furthermore the so called Qblocks interdigitate bringing two or more β-sheets into close lateral proximity. This favourable steric effect has been reported in other well-studied amyloidogenic systems.[22]

Sikorski and Atkins suggest that in general where Qblocks are present, they may form a nucleation site with the potential to encourage aggregation and β-sheet formation in other segments of proteins,[31] furthermore could a Qblock be used as a template for the assembly of other functional proteins? To this effect a Qblock template could be conjugated to a biologically active molecule such that upon self-assembly a functionalised fibre is formed whilst retaining the structural advantages of a self-assembled system. Indeed it has been demonstrated that a functional conducting amyloid can be made by conjugating cytochrome b_{562} to an SRC homology 3 (SH3) domain. The SH3 domain, a small natively globular protein consisting of β-strands in a β-barrel fold, is later promoted to form the amyloid core by incubating in the appropriate conditions.[1] Baldwin *et al.* show that the cytochrome is displayed on the amyloid fibril at high density and can bind porphyrin successfully suggesting it is functional. The high density infers the potential for rapid electron transfer making such a system a good candidate for a man-made biological conductor.

The extent to which so many different proteins of unique primary sequence can adopt the same molecular architecture is remarkable. Clearly aberrant self-assembly and aggregation is a pathway shared by many unfolded polypeptides, but it is also a pathway used by particular systems in a functional role. Elucidation of the structure of these assemblies is not straightforward but doing so promises to deliver insights not only into possible therapeutics but also to inspire a plethora of biomaterials.

De-novo designed fibrillating peptides

Self-assembling short peptides not always related to any native protein sequence have been engineered based on a set of design principles—side chains attracted *via* hydrophobic forces, aromatic residues for π–π interactions, electrostatic interactions *via* oppositely charged residues and alternating hydrophilic-hydrophobic residues to make the peptide molecule asymmetrical in its attraction to water.[32] The resulting designer peptides successfully assemble in water from monomer into self-supporting β-sheet rich tapes producing thermostable gels. Further work has demonstrated the ability to design gelating self-assembling materials that are tunably responsive to external stimuli. A self-assembling hydrogel can be formed using amino acids that have a high propensity for forming β-sheets and including a sequence designed to form a type-II β-hairpin.[33] The use of high β-sheet propensity amino acids and the positioning of the β-hairpin signature affords an impressive amount of control and potential for modifying material properties but the choice of the amino acid sequence is key to factors stabilising these structures. A repeating $[K\text{-}V]_n$ motif produces an amphiphilic molecule, resulting in increased lateral association that probably results in the formation of highly entangled networks of fibres that can trap water and give rise to a hydrogel.

Derivations of the basic peptide design have rational effects on material properties whereby the assembly can be controlled by altering pH and temperature. The pH at which assembly occurs can be modulated by modifications to the hydrophilic side of the molecule, mutations to hydrophobic residues result in the removal of charge–charge repulsion and thus self-assembly occurs at lower and importantly more physiological pHs. Conversely the temperature at which assembly occurs can be modulated by modifications to the hydrophobic side of the molecule, whereby increasing temperature results in the desolvation of hydrophobic residues and consequently self-assembly. Thus mutations to less hydrophobic residues on the hydrophobic side of the peptide result in a system that self-assembles at higher temperatures due to the less readily desolvating residues.[3]

Di-peptide hydrogels

As discussed previously, there are a number of self-assembling β-sheet forming short peptides that have been reported in the literature but a step away from these generally biologically inspired systems introduces even shorter peptide systems. The conjugation of large aromatic groups allows the length of the peptide to be reduced whilst retaining assembly potential.[34,35] These small molecules, called low molecular weight gelators (LMWG), are typically observed to form soft hydrogels.

Two such systems involve the conjugation of FMOC or naphthalene to a di-peptide. As with previously designed de-novo systems their assembly can be controlled by external factors such as temperature or pH. The naphthalene conjugated system forms soft hydrogels at low pH and so its assembly can be controlled by reducing the pH, however the FMOC-diphenylalanine system is unique in that it forms similar hydrogels at physiological pH,[34] making it an attractive system to investigate given the potential for use in humans *i.e.* drug release or tissue engineering. Detailed spectroscopic and X-ray diffraction investigations have shown both systems adopt a β-sheet molecular architecture with an emphasis placed on π–π stacking in the FMOC system. X-Ray diffraction data has inferred similarities between the naphthalene dipeptide systems, amyloidogenic systems and the FMOC system[35] revealing insights into the assembly process. It would seem that the large aromatic groups facilitate a general aggregation

of the monomers *via* a hydrophobic collapse and subsequent establishment of hydrogen bonding produces order and directionality in the system such that it forms fibres.

The rational design of these systems to date has been limited. However through their characterisation and our increased understanding of the driving forces behind their assembly we will be able to make informed changes to the molecular characteristics of the monomers to predictably produce desired characteristics in the macromolecular assemblies.

Self-assembling coiled-coils

Alpha-helical coiled-coils formed by a repetitive seven residue structural motif are found in many biological systems such as α-keratin and myosin and the rules that govern the final assembly are well understood. Twenty-eight residue peptides made up of the heptad consensus "abcdefg" have been designed to assemble *via* sticky ends to form a repeating heterodimeric unit that produces filaments with lengths in the order of several hundred micrometres that bundle to form matured fibres[36] (Fig. 1c). The fibres are around 50 nm in diameter, and with pronounced striations as shown in Fig. 2c. These assemblies are referred to as SAFs (self-assembling fibres) and the principles of their design are to align interacting amino acids in the positions that are responsible for dimer interface at positions a,d,e and g.[36]

In line with the desire to be able to modify the molecular features of a monomer to change the properties of the final self-assembled material, variations of the basic SAF design have been engineered to produce hydrogels. By maintaining the residues responsible for dimerisation interface but modifying the remaining residues specifically, lateral fibre interactions can be modulated to be based on weak hydrophobic and hydrogen bonding interactions. These changes produce a new class of hydrogelating SAFs.[2]

Structural determination of self-assembled systems

Structure determination of fibrous molecules using conventional techniques of X-ray crystallography and solution nuclear magnetic resonance is hampered due to the inherent heterogeneity and insolubility of the final assemblies. Recently, several short amyloidogenic peptides have been successfully crystallised and these are a valuable exception.[22] However some questions remain as to whether the interactions observed in the crystal structures are the same as those that stabilise the fibrils.[37,38] Traditionally, the structures of fibrous molecules have been investigated using X-ray fibre diffraction yielding high-resolution information and mechanistic insights; a particularly famous example being the structure of DNA by Franklin, Crick and Watson.[39] This method gives information about the long range crystalline or semi-crystalline order that is present along the fibre axis. The fibre sample is prepared in order to align individual fibrils on one axis such that all are laterally associated. Alignment of fibrils may be achieved in a variety of ways as shown in Fig. 3. The stretch frame method and a derivative using wax-tipped capillaries shown in Fig. 3a has been used extensively for alignment of amyloid fibrils[21,40–42] and coiled-coil self-assembled fibres.[2,36] To obtain more information from fibrillar assemblies or if alignment has

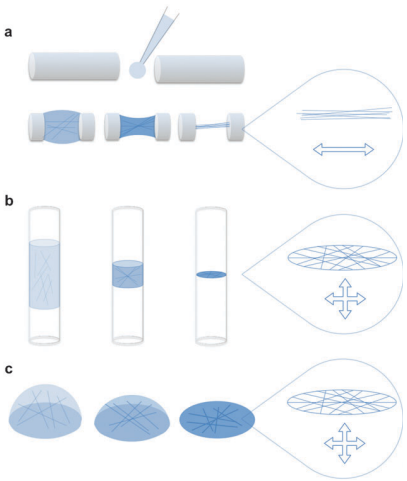


Fig. 3 Methods of dehydration for the alignment of fibrous proteins, corresponding fibril axes are shown to the right of illustrations. A solution of fibrils may be suspended between two sealed capillaries to create a bundle of fibrils aligned parallel to one another (a) or suspended inside a capillary and dried to a disc with fibrils aligned at random orientations to one other but on the same plane (b). Finally, a similar texture may be made by drying a drop to a mat or film (c).

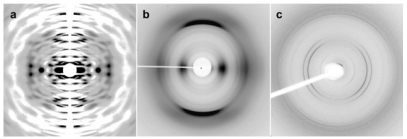


Fig. 4 X-Ray fibre diffraction patterns of the mixed secondary globular structure of actin[48] (a), cross-β laminated sheets of Aβ(1–40) amyloid fibrils[43] (b) and α-helical coiled coils of self-assembling filaments (SAFs)[36] (c). All fibre axes are vertical.

proven unsuccessful, additional methods may be attempted. Fig. 3b and c illustrate alterative methods of alignment that result in the production of thin films where the rotational axes are more complicated and must be taken into account during analysis.[43–46]

X-Ray diffraction patterns collected from a fibre bundle have two axes, referred to as the equator and meridian, and are a consequence of the directionality in the sample. The repetition of subunits along the fibril axis gives rise to an inherent crystalline order in one direction, with corresponding reflections observed on the meridian. Order on the axes perpendicular to the fibril axis are combined and radially averaged, with corresponding reflections observed on the

equator. The rotational arrangement of fibrils relative to the vertical fibre axis results in the diffraction signal spots forming arcs. Fig. 4 illustrates the typical types of patterns obtained from this technique with varying degrees of radial averaging due to varying alignment.

Data processing for X-ray fibre diffraction patterns is described in detail elsewhere[45,47] but simple analysis and interpretation may be applied to various fibrillar samples (with a fibre texture) if it is considered that reflections on the meridian relate to unit cell dimensions along the fibril axis and reflections on the equator are a mixture relating to the perpendicular unit cell axes. Given likely unit cell dimensions, peptides can be modelled and the diffraction pattern can be calculated from these structures to allow comparison between calculated and experimental patterns to validate possible models.[5,21,36] In line with this method, detailed interpretations of Aβ(11–25),[43] collagen,[44] F-actin[48,49] and the self-assembling SAFs[36] have been reported in the literature.

Structure determination using X-ray fibre diffraction can be complemented using a range of other structural, imaging and biophysical techniques. Advances in solid state NMR (ssNMR) have enabled detailed structures of fibrous proteins, in particular amyloid-like fibrils, to be elucidated[50,51] and ssNMR data can also be combined with the information gathered from X-ray fibre diffraction to generate possible model structures.[52] Additionally known crystallographic structures of subunits may be modelled against X-ray diffraction data to produce high-resolution models. Examples of this approach include muscle[41] and actin[48] (Fig. 1a). Electron microscopy and atomic force microscopy can contribute important information regarding the diameters of filaments, as well as repeating units or helical twists that may be visible under the microscope. This information can be combined with X-ray fibre diffraction analysis to aid the generation of model structures[36,53] (for review see ref. 54). Additionally cryo-electron microscopy studies have facilitated the construction of low resolution three-dimensional structures of fibres,[55] a combination of high resolution information with these low resolution models can reveal how short range order affects long range order. Indeed recently, significant advances have been made in understanding the amyloidogenic proteins using small angle X-ray scattering analysis[56] producing low resolution models that could be combined with high resolution models. This technique is also performed in solution and so has yielded useful information about the assembly process. Other methods for studying final structure and the assembly process include circular and linear dichroism. Due to the chirality of amino acids, a polypeptide chain will exhibit different optical behaviours in circularly polarised light depending on their secondary structure.[57] Linearly polarised light is also seen to have unique optical effects on proteins but only when on the correct axis. This phenomenon can be problematic when trying to study fibrillar proteins by CD if the sample has a tendency to spontaneously align and so its contributions must be eliminated to ensure good analysis of CD spectra. However if the appropriate considerations are made and instrumentation is available then the interactions of linearly polarised light with fibrillar proteins can reveal orientation information about chromophores in the sample.[38,58]

Advances in understanding the structures of self-assembling peptides using X-ray fibre diffraction

In the final section of this review, we will examine the information that has been obtained using X-ray fibre diffraction on fibrous assemblies concentrating on three examples: fibrous actin; amyloid and coiled-coil SAFs.

Actin

Actin exists in two forms: G-actin, the globular form and F-actin, a polymerised fibrillar form.[11] It was shown in early work that an increase in salt concentration drives a transition from G-actin to F-actin and so the two forms could be isolated for structural studies.[59] Methods for aligning actin fibres had already been published in detail[60] and X-ray fibre diffraction patterns successfully collected. Later the high resolution crystal structure of G-actin in the ATP and ADP bound forms were determined in 1990 by Kabsch *et al.*[61] At the same time a model for F-actin was proposed. This was based on the structural determination of F-actin by TEM, the low-resolution constraints of this model were used to fit the high resolution atomic co-ordinates of G-actin into an F-actin model.[49] Close comparisons between the calculated diffraction pattern and the experimental fibre diffraction pattern confirmed the models validity and led to the general adoption of this model as representative of F-actin.

Subsequently, a reiterative computational approach has been used in an attempt to better match the experimental fibre diffraction and represent the changes that might occur in the actin fibrillar transition. Refinement of the F-actin model and calculation of a diffraction pattern almost completely reproduces the experimental fibre diffraction pattern.[62] This represents a multi-technique approach that reflects the need for corroborative data in X-ray fibre diffraction analysis. This model has been accepted and used successfully to describe additional molecular features of actin in muscle fibres.[63] Models produced in this manner may continually be refined and improved but most significantly in the advent of technical advances in fibre diffraction and sample preparation (magnetic alignment and fibril length control) new data at higher resolution has recently been reported, shown in Fig. 4a. Combined with computational refinement the new data has been used to generate a new model of F-actin from the original G-actin crystallographic model.[48] The authors note differences in the new F-actin model that result in greater stabilising contacts between monomers but an inherent instability associated with the monomer conformations themselves, perhaps giving structural insights into actin dynamics.

Amyloid

Amyloid fibrils formed by Bence Jones immunoglobulin fragments and isolated from a patient suffering from multiple myeloma, were observed to have a cross-β diffraction pattern by X-ray fibre diffraction.[64] This pattern was shown to have similar details to those observed for cross-β silk from the Lacewing fly egg-stalk.[65] An example of the cross-β pattern from X-ray fibre diffraction is shown in Fig. 4b, and this pattern is typical of that collected from amyloid fibrils formed from many different amyloidogenic proteins. Interpretation of these patterns gave rise to a structure in which β-strands run perpendicular to the fibre axis, and the β-sheets are hydrogen bonded, 4.76 Å apart along the length of the fibre. Several sheets associate, *via* side chain interactions giving rise to a diffuse and variable equatorial reflection.[64] This conclusion was supported by the observation that the silk pattern gave a 5 Å reflection on the equator, compared to the 10 Å reflection observed for amyloid, consistent with differing sizes of side chains (silk has a repetitive sequence containing the small residues A and G).[65] Diffraction from a range of amyloid fibrils extracted from diseased tissue and formed by transthyretin, lysozyme, immunoglobulin light chain as well as those formed by short synthetic peptides all give a similar cross-β diffraction pattern reinforcing the view that amyloid fibrils formed from different proteins and peptides share a common core structure.[5]

The general view of the cross-β core for amyloid has been challenged and a β-helix based structure has been proposed.[66,67] However, careful analysis of a range of structural models using analysis of experimental and calculated fibre diffraction supports the cross-β structure as being the core structure for amyloid fibrils.[5] It does appear however that some features of a β-solenoid structure may be supported by the cross-β diffraction pattern. Recently, the β-solenoid structure of the Het-S prion amyloid-like fibrils has been elucidated by ssNMR[51] and calculation of the fibre diffraction pattern from this model structure shows some features consistent with cross-β patterns. This likely arises from the close packing of the β-sheets within the solenoid structure, which differs very little from the general model for cross-β.[5] Recently, interesting studies of *ex vivo* extracted prion filaments have revealed a different diffraction pattern that requires the smallest repeating unit to be larger than the separation of hydrogen bonded β-strands[67] and this may reveal differences in the structures of prion rods compared to general amyloid fibrils.

Established model structures for the Alzheimer's amyloid fibril are made up of stacks of parallel β-sheets connected *via* a β-bend to form two β-sheets (Fig. 1b). The sheets are held together by close association of the side chains similar to those suggested for the steric zippers determined from crystallography of short amyloidogenic peptides.[22] X-Ray fibre diffraction combined with electron diffraction analysis of fibrous crystals formed by a 12mer designed peptide gave a cross-β pattern that had enough features to enable the determination of a unit cell followed by modelling and diffraction calculation. This gave rise to a model structure revealing that association of the aromatic phenylalanine residues between and along the β-sheets may stabilise the structures.[21]

X-Ray diffraction has been used to support the structural analysis of amyloid fibrils formed by a fragment of the Diabetes type 2 peptide islet amyloid polypeptide (IAPP) residues 20–29. This peptide formed highly ordered, semi crystalline fibrils that were studied using ssNMR. Calculation of the diffraction patterns from models prepared from ssNMR data were compared to experimental diffraction patterns to arrive at conclusions regarding the probable arrangement of the peptide within the fibrils.

Designer peptides have been engineered to switch between coiled coil and β-sheet structure upon a change in conditions such as temperature.[53,68] X-Ray fibre diffraction from β-sheet rich amyloid fibrils formed from the ccβMet peptide enabled the structural modelling for a β-hairpin structure that was consistent with other structural techniques including scanning transmission electron microscopy and atomic force microscopy.[53]

Coiled coil self-assembling fibres

Two complementary leucine zipper-based peptides were engineered with sticky ends to promote longitudinal assembly.[36] These peptides assembled to form 40–50 nm wide fibres, many microns long and electron microscopy showed evenly spaced striations arising from the association of the 4.2 nm long coiled-coil dimers. X-Ray fibre diffraction from partially dried, aligned fibre bundles revealed oriented diffraction signals consistent with the expected repeats for a heptad repeat coiled coil at 1.036 nm shown in Fig. 4c. The observation of this repeat on the meridian of the pattern is consistent with the arrangement of the coiled coil peptides parallel to the fibre axis. A series of equatorial diffraction signals arise from the packing of coiled coil dimers and could be indexed to support hexagonal packing of the cylinders with dimensions of 1.824 nm. Modelling of the coiled coil peptides into a packing arrangement gave rise to an overall structural model (Fig. 1c) from which a diffraction pattern could be calculated. The calculated pattern matched extremely well to the experimental data, suggesting that the model was consistent with the experimental data.

Further design of the SAFs gave rise to a new generation that formed hydrogels and named hSAFs. A series of amino-acid substitutions into the original sequence enabled the formation of fibres with differing properties, particularly in terms of their lateral association.[2] The positions that are exposed on the coiled coil surface (b, c and f) were either substituted for all A to promote weak hydrophobic interactions or all Q for its propensity to hydrogen bond giving $hSAF_{AAA}$ and $hSAF_{QQQ}$.[2] X-Ray fibre diffraction from $hSAF_{AAA}$ and $hSAF_{QQQ}$ revealed that the essential elements of the coiled coil assembly remained similar to the original design, although the diffraction patterns were significantly less oriented indicating a poorer crystalline arrangement within the samples. This is consistent with the formation of narrower filaments that are tangled to form hydrogels. The packing differed from the original (18.2 Å) giving a tighter packing for the A substitution (17.3 Å) and a wider packing (21.5 Å) for the Q variant.

Conclusions

Clearly self-assembly is a biologically conserved mechanism and useful way of producing macromolecular scale structures. Large proteins may self-assemble intra and extra-cellularly in prokaryotic and eukaryotic cells. The extent to which so many proteins of unique primary sequence and secondary structure undergo this process is remarkable. They may be used alone or as a composite to complement the inherent strength of hierarchical assemblies. Conversely, aberrant self-assembly is associated with a number of diseases, although progress in understanding the assembly mechanism could reveal therapeutic insights. Further it is being demonstrated that the template

This journal is © The Royal Society of Chemistry 2010

that leads to disease has promising applications for materials science.

From natural to designer systems these full-length proteins are inspiring a range of short peptide systems where the lengths can be reduced from around ten amino acid residues to two by way of aromatic conjugations. Design rules are being developed and used based on overall physicochemical properties and in tandem the investigation of these properties finds some correlation to aggregation in natural systems. Position-specificity is already well applied in α-helical coiled coils and analyses of β-sheet structures in time should lead to finer control over assembly as well as revealing more detail about the complex natural systems.

The study of fibrillar assemblies has generally been difficult using traditional structural determination methods. A multi-technique approach is required to fully characterise this kind of system but to date X-ray fibre diffraction has certainly played a central role. Sample preparation is key and following from this the degree of fibre alignment directly impacts on the quality of data obtained. Any solution of fibres may be aligned simply based on the unidirectionality of the system but the degree of alignment is sample dependent and so it is essential to try different alignment methods to obtain the most information.

The extraction of structural information from fibre diffraction patterns is not an entirely standardised procedure and this is reflected in the variety of methodologies in analysis and model building. This only further illustrates the necessity for a multi-technique approach. Generally, high-resolution structures (crystallography or ssNMR) may be built into low-resolution models (TEM or SAXS) followed by fibre diffraction calculation; medium-resolution models may be built (fibre diffraction) and other studies infer secondary structure and side chain placement (CD or LD) followed by fibre diffraction calculation or a combination of these two can be used with reiterative computational refinement to lead to successful modelling/interpretation. Ultimately any interpretive model building process is concluded with a comparison of a calculated with the experimental fibre diffraction pattern.

Studies in recent decades have advanced our awareness and understanding of a huge breadth of fibrillar protein systems. The technical challenge in their study has encouraged the development of methodologies and given insight into natural and disease processes, developed our understanding of protein engineering and begun to deliver a range of promising nanomaterials.

Acknowledgements

Transmission electron micrographs of actin were kindly provided by Dr Carolyn Moores (University of London, Birbeck College) and coiled coil SAFs by Prof. Dek Woolfson (University of Bristol). Aβ-amyloid was generously provided by Prof. Dr Kerensa Broersen and imaged by Kyle Morris (University of Sussex). The fibre diffraction pattern of actin was kindly supplied by Dr Toshiro Oda (RIKEN SPring-8 Center) and reproduced with permission.

LS acknowledges funding from BBSRC and Alzheimer's Research Trust.

Notes and references

1 A. J. Baldwin, R. Bader, J. Christodoulou, C. E. MacPhee, C. M. Dobson and P. D. Barker, *J. Am. Chem. Soc.*, 2006, **128**, 2162–2163.
2 E. F. Banwell, E. S. Abelardo, D. J. Adams, M. A. Birchall, A. Corrigan, A. M. Donald, M. Kirkland, L. C. Serpell, M. F. Butler and D. N. Woolfson, *Nat. Mater.*, 2009, **8**, 596–600.
3 R. V. Rughani and J. P. Schneider, *MRS Bulletin*, 2008, **33**, 530–535.
4 C. Vendrely and T. Scheibel, *Macromol. Biosci.*, 2007, **7**, 401–409.
5 T. R. Jahn, O. S. Makin, K. L. Morris, K. E. Marshall, P. Tian, P. Sikorski and L. C. Serpell, *J. Mol. Biol.*, 2010, **395**, 717–727.
6 G. M. Whitesides, J. P. Mathias and C. T. Seto, *Science*, 1991, **254**, 1312–1319.
7 D. Sept, N. A. Baker and J. A. McCammon, *Protein Sci.*, 2003, **12**, 2257–2261.
8 J. B. Olmsted and G. G. Borisy, *Annu. Rev. Biochem.*, 1973, **42**, 507–540.
9 C. H. Lin and P. Forscher, *Neuron*, 1995, **14**, 763–771.
10 E. Nogales, in *Annual Review of Biophysics and Biomolecular Structure*, ed. R. M. Stroud, W. K. Olson and M. P. Sheetz, Annual Reviews {a}, 2001, pp. 277–302.
11 T. D. Pollard, *Curr. Opin. Cell Biol.*, 1990, **2**, 33–40.
12 X. Wang, N. D. Hammer and M. R. Chapman, *J. Biol. Chem.*, 2008, **283**, 21530–21539.
13 X. Wang and M. R. Chapman, *J. Mol. Biol.*, 2008, **380**, 570–580.
14 A. Sponner, W. Vater, S. Monajembashi, E. Unger, F. Grosse and K. Weisshart, *PLoS One*, 2007, **2**, e998.
15 S. Keten and M. J. Buehler, *Comput. Methods Appl. Mech. Eng.*, 2008, **197**, 3203–3214.
16 S. Rammensee, U. Slotta, T. Scheibel and A. R. Bausch, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 6590–6595.
17 J. G. Hardy and T. R. Scheibel, *Biochem. Soc. Trans.*, 2009, **37**, 677–681.
18 J. A. Hardy and G. A. Higgins, *Science*, 1992, **256**, 184–185.
19 C. M. Dobson, *Nature*, 2003, **426**, 884–890.
20 G. B. Irvine, O. M. El-Agnaf, G. M. Shankar and D. M. Walsh, *Mol. Med.*, 2008, **14**, 451–464.
21 O. S. Makin, E. Atkins, P. Sikorski, J. Johansson and L. C. Serpell, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 315–320.
22 M. R. Sawaya, S. Sambashivan, R. Nelson, M. I. Ivanova, S. A. Sievers, M. I. Apostol, M. J. Thompson, M. Balbirnie, J. J. Wiltzius, H. T. McFarlane, A. O. Madsen, C. Riekel and D. Eisenberg, *Nature*, 2007, **447**, 453–457.
23 A. P. Pawar, K. F. Dubay, J. Zurdo, F. Chiti, M. Vendruscolo and C. M. Dobson, *J. Mol. Biol.*, 2005, **350**, 379–392.
24 G. G. Tartaglia, A. Cavalli, R. Pellarin and A. Caflisch, *Protein Sci.*, 2005, **14**, 2723–2734.
25 M. López de la Paz and L. Serrano, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 87–92.
26 F. Rousseau, J. Schymkowitz and L. Serrano, *Curr. Opin. Struct. Biol.*, 2006, **16**, 118–126.
27 S. Zibaee, O. S. Makin, M. Goedert and L. C. Serpell, *Protein Sci.*, 2007, **16**, 906–918.
28 M. J. Thompson, S. A. Sievers, J. Karanicolas, M. I. Ivanova, D. Baker and D. Eisenberg, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 4074–4078.
29 S. Maurer-Stroh, M. Debulpaep, N. Kuemmerer, M. L. de la Paz, I. C. Martins, J. Reumers, K. L. Morris, A. Copland, L. Serpell, L. Serrano, J. W. Schymkowitz and F. Rousseau, *Nat. Methods*, 2010, **7**, 237–242.
30 C. P. Jaroniec, C. E. MacPhee, V. S. Bajaj, M. T. McMahon, C. M. Dobson and R. G. Griffin, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 711–716.
31 P. Sikorski and E. Atkins, *Biomacromolecules*, 2005, **6**, 425–432.
32 A. Aggeli, M. Bell, N. Boden, J. N. Keen, T. C. B. McLeish, I. Nyrkova, S. E. Radford and A. Semenov, *J. Mater. Chem.*, 1997, **7**, 1135–1145.
33 R. V. Rughani, D. A. Salick, M. S. Lamm, T. Yucel, D. J. Pochan and J. P. Schneider, *Biomacromolecules*, 2009, **10**, 1295–1304.
34 A. M. Smith, R. J. Williams, C. Tang, P. Coppo, R. F. Collins, M. L. Turner, A. Saiani and R. V. Ulijn, *Adv. Mater.*, 2008, **20**, 37–41.

Downloaded by University of Sussex on 08 September 2011
Published on 29 July 2010 on http://pubs.rsc.org | doi:10.1039/B919453N

35 L. Chen, K. Morris, A. Laybourn, D. Elias, M. R. Hicks, A. Rodger, L. Serpell and D. J. Adams, *Langmuir*, 2010, **26**, 5232–5242.
36 D. Papapostolou, A. M. Smith, E. D. Atkins, S. J. Oliver, M. G. Ryadnov, L. C. Serpell and D. N. Woolfson, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 10853–10858.
37 P. C. van der Wel, J. R. Lewandowski and R. G. Griffin, *J. Am. Chem. Soc.*, 2007, **129**, 5117–5130.
38 K. E. Marshall, M. R. Hicks, T. L. Williams, S. V. Hoffmann, A. Rodger, T. R. Dafforn and L. C. Serpell, *Biophys. J.*, 2010, **98**, 330–338.
39 M. Coles, W. Bicknell, A. A. Watson, D. P. Fairlie and D. J. Craik, *Biochemistry*, 1998, **37**, 11064–11077.
40 O. S. Makin and L. C. Serpell, *J. Mol. Biol.*, 2004, **335**, 1279–1288.
41 A. M. Squires, G. L. Devlin, S. L. Gras, A. K. Tickler, C. E. MacPhee and C. M. Dobson, *J. Am. Chem. Soc.*, 2006, **128**, 11738–11739.
42 T. R. Serio, A. G. Cashikar, A. S. Kowal, G. J. Sawicki, J. J. Moslehi, L. Serpell, M. F. Arnsdorf and S. L. Lindquist, *Science*, 2000, **289**, 1317–1321.
43 P. Sikorski, E. D. T. Atkins and L. C. Serpell, *Structure*, 2003, **11**, 915–926.
44 P. Sikorski, M. Wada, L. Heux, H. Shintani and B. T. Stokke, *Macromolecules*, 2004, **37**, 4547–4553.
45 O. S. Makin and L. C. Serpell, in *Amyloid Proteins. Methods and Protocols*, ed. E. M. Sigurdsson, Humana Press, 2005, vol. 299, pp. 67–80.
46 J. M. Wang, S. Gulich, C. Bradford, M. Ramirez-Alvarodo and L. Regan, *Structure*, 2005, **13**, 1279–1288.
47 G. Rajkumar, H. A. Al-Khayat, F. Eakins, C. Knupp and J. M. Squire, *J. Appl. Crystallogr.*, 2007, **40**, 178–184.
48 T. Oda, M. Iwasa, T. Aihara, Y. Maeda and A. Narita, *Nature*, 2009, **457**, 441–445.
49 K. C. Holmes, D. Popp, W. Gebhard and W. Kabsch, *Nature*, 1990, **347**, 44–49.
50 R. Tycko, *Curr. Opin. Struct. Biol.*, 2004, **14**, 96–103.
51 C. Wasmer, A. Lange, H. Van Melckebeke, A. B. Siemer, R. Riek and B. H. Meier, *Science*, 2008, **319**, 1523–1526.
52 J. Madine, E. Jack, P. G. Stockley, S. E. Radford, L. C. Serpell and D. A. Middleton, *J. Am. Chem. Soc.*, 2008, **130**, 14990–15001.
53 M. O. Steinmetz, Z. Gattin, R. Verel, B. Ciani, T. Stromer, J. M. Green, P. Tittmann, C. Schulze-Briese, H. Gross, W. F. van Gunsteren, B. H. Meier, L. C. Serpell, S. A. Muller and R. A. Kammerer, *J. Mol. Biol.*, 2008, **376**, 898–912.
54 T. Stromer and L. C. Serpell, *Microsc. Res. Tech.*, 2005, **67**, 210–217.
55 J. L. Jimenez, E. J. Nettleton, M. Bouchard, C. V. Robinson, C. M. Dobson and H. R. Saibil, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 9196–9201.
56 B. Vestergaard, M. Groenning, M. Roessle, J. S. Kastrup, M. van de Weert, J. M. Flink, S. Frokjaer, M. Gajhede and D. I. Svergun, *PLoS Biol.*, 2007, **5**, e134.
57 S. M. Kelly, T. J. Jess and N. C. Price, *Biochim. Biophys. Acta, Proteins Proteomics*, 2005, **1751**, 119–139.
58 T. R. Dafforn, J. Rajendra, D. J. Halsall, L. C. Serpell and A. Rodger, *Biophys. J.*, 2004, **86**, 404–410.
59 F. B. Straub, *Studies, University of Szeged*, 1942, **II**, 1–15.
60 D. Popp, V. V. Lednev and W. Jahn, *J. Mol. Biol.*, 1987, **197**, 679–684.
61 W. Kabsch, H. G. Mannherz, D. Suck, E. F. Pai and K. C. Holmes, *Nature*, 1990, **347**, 37–44.
62 M. Lorenz, D. Popp and K. C. Holmes, *J. Mol. Biol.*, 1993, **234**, 826–836.
63 P. Vibert, R. Craig and W. Lehman, *J. Mol. Biol.*, 1997, **266**, 8–14.
64 E. D. Eanes and G. G. Glenner, *J. Histochem. Cytochem.*, 1968, **16**, 673–677.
65 A. J. Geddes, K. D. Parker, E. D. Atkins and E. Beighton, *J. Mol. Biol.*, 1968, **32**, 343–358.
66 M. F. Perutz, J. T. Finch, J. Berriman and A. Lesk, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 5591–5595.
67 H. Wille, W. Bian, M. McDonald, A. Kendall, D. W. Colby, L. Bloch, J. Ollesch, A. L. Borovinskiy, F. E. Cohen, S. B. Prusiner and G. Stubbs, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 16990–16995.
68 B. Ciani, E. G. Hutchinson, R. B. Sessions and D. N. Woolfson, *J. Biol. Chem.*, 2002, **277**, 10150–10155.
69 T. Luhrs, C. Ritter, M. Adrian, D. Riek-Loher, B. Bohrmann, H. Dobeli, D. Schubert and R. Riek, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 17342–17347.
70 W. L. DeLano, *The PyMOL Molecular Graphics System*, http://www.pymol.org.

The delicate balance between gelation and crystallisation: structural and computational investigations†

Dave J. Adams,*[a] Kyle Morris,[b] Lin Chen,[a] Louise C. Serpell,[b] John Bacsa[a] and Graeme M. Day*[c]

Received 21st May 2010, Accepted 10th June 2010
DOI: 10.1039/c0sm00409j

Predicting the ability of low molecular weight molecules to form hydrogels is difficult. Here, we have examined the self-assembly behavior of two chemically and structurally similar functionalized dipeptides, one of which is found to form a meta-stable hydrogel (**1**) and the other forming a crystalline solid (**2**). To investigate the reasons for these differences, we have employed computational methods to explore the crystal energy landscapes of the two molecules and examined differences in their preferred packing arrangements. We show that this method accurately predicts the packing for the crystalline solid, **2**. Furthermore, the predictions for the gel-former **1** suggest that one-dimensional hydrogen-bonding arranged into tightly coiled molecular columns is a preferred mode of packing for this system, but is unfavorable for **2**. The different tendencies of forming these columns could provide an explanation for the different behavior of the two molecules and demonstrate that this approach could be useful for the future predictable design of low molecular weight gelators.

Introduction

There is currently significant interest in the assembly of low molecular weight molecules to give gels.[1,2] Gels are formed *via* the formation of a fibrous network, resulting from the assembly of the gelator molecules through uni-directional non-covalent forces such as hydrogen-bonding, π–π stacking and van der Waals interactions.[3] This approach has been shown to be successful for a number of very different systems.[1,2,4] For example, hydrogels can be prepared from β-sheet forming oligopeptides,[5–11] α-helix forming oligopeptides,[12] *N*-functionalised dipeptides,[13–22] 1,3,5-trisamide-cyclohexanes[23] in addition to functionalized sugars.[24] This enormous structural diversity coupled with different triggers by which assembly occurs (for example pH, temperature and enzymatic action) means that there are limited correlations between the molecular structure and the intermolecular interactions leading to assembly.[24,25] Indeed, there is a lack of predictability in the design of the gelators. Whilst many gelators are known, serendipity still plays a role in their discovery.[26]

A balance between crystallization and solubilization is needed to form three-dimensional networks of fibers. It is normally thought that self-complementary and unidirectional non-covalent interactions drive one-dimensional growth leading to fibers.[26] An important question is the relationship between the 3-dimensional packing of the gelator molecules in a crystal and the (possibly) less ordered aggregates leading to gels. Indeed, the crystal engineering approach towards the design of low molecular weight hydrogelators has demonstrated that there is often a link between the type of packing observed in a crystal structure of a molecule and its ability to act as a gelator.[26] On the other hand, equilibrium kinetics predicts that the stable thermodynamic phase will be crystalline, with the gel phase representing a kinetically trapped phase. This has been demonstrated by a gel phase forming a crystalline phase over time. For example, a study on a sugar–biphenyl gelator showed that a metastable gel could be formed in a water/1,4-dioxane mixture that collapsed giving a crystalline state. The wide-angle X-ray diffraction of the crystalline state was found to be the same as that of a single crystal grown from acetonitrile. However, the diffraction from a xerogel was very different and suggested a lamellar packing.[27] Elsewhere, it has been shown for a two component gelator system that fibers and crystals are formed in a step-wise fashion, meaning that it is difficult to directly link the final crystal structure and the structure in the fiber phase.[28,29] It has also been shown by powder X-ray diffraction that the structure in the gel fibers formed from a cholesteryl-derived organogelator was not the same as that obtained on crystallization, demonstrating that packing in the gel and the crystalline phases are not always correlated.[30]

It is clear that a simple method able to predict the packing of potential hydrogelators and hence whether hydrogelation will occur would revolutionize this field. Here, we describe a study that is intended as a first step towards this difficult goal. We have examined the self-assembly behavior of two chemically and structurally similar functionalized dipeptides. One of these is found to form a meta-stable hydrogel, from which crystals of suitable quality for X-ray diffraction can be grown under the correct conditions. The other forms a crystalline solid directly. To investigate these differences in behavior, we have employed computational methods to explore the crystal energy landscapes of the two molecules. We show that the calculations accurately predict the packing for the crystalline solid as one of the low energy possibilities and that the two molecules are predicted to have different hydrogen bonding and packing preferences. There is a greater preference for unidirectional hydrogen-bonding in the gel-former and tightly coiled columns of hydrogen-bonded molecules are favored for this system.

[a]*Department of Chemistry, University of Liverpool, Crown Street, Liverpool, L69 7ZD, UK. E-mail: d.j.adams@liverpool.ac.uk*
[b]*Department of Chemistry and Biochemistry, School of Life Sciences, University of Sussex, Falmer, BN1 9QG, UK*
[c]*Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, UK. E-mail: gmd27@cam.ac.uk*

† Electronic supplementary information (ESI) available: Full synthesis and characterization for dipeptides **1** and **2** and all intermediates. CCDC reference numbers 764715 and 764714. For ESI and crystallographic data in CIF or other electronic format see DOI: 10.1039/c0sm00409j

View Online

Downloaded by University of Sussex on 17 February 2011
Published on 07 July 2010 on http://pubs.rsc.org | doi:10.1039/C0SM00409J

Experimental

Materials

All chemical reagents and solvents were purchased from Sigma Aldrich and used without further purification. Distilled water was used throughout. The synthesis and characterization of the dipeptides is described in the ESI†.

NMR

^{1}H NMR spectra were recorded at 400.13 MHz using a Bruker Avance 400 NMR spectrometer. ^{13}C NMR spectra were recorded at 100.6 MHz.

Scanning electron microscopy

High resolution imaging of the polymer morphology was achieved using a Hitachi S-4800 cold Field Emission Scanning Electron Microscope (FE-SEM). Samples of hydrogels were placed on a 15 mm Hitachi M4 aluminium stub using an adhesive high purity carbon tab and allowed to air dry. The samples were then coated with a 2 nm layer of gold using an Emitech K550X automated sputter coater. The FE-SEM measurement scale bar was calibrated using certified SIRA calibration standards. Imaging was conducted at a working distance of 8 mm and a working voltage of 3 kV using a mix of upper and lower secondary electron detectors.

Single crystal X-ray diffraction for dipeptide 1

$C_{17}H_{18}N_2O_5$, $M_w = 330.33$, orthorhombic, space group $P2_12_12_1$, $a = 9.9170(10)$ Å, $b = 11.1123(11)$ Å, $c = 12.511(2)$ Å, $V = 3112.8(6)$ Å^3, $Z = 8$, $D_{calc} = 1.410$ g cm^{-3}. Exposure times of 30 s per frame and scan widths of 0.3° were used throughout the data collection. The data collection was performed using three ω scans yielding data in the range 1.44 to 27.10° with an average completeness of 99%.

Single crystal X-ray diffraction for dipeptide 2

$C_{17}H_{18}N_2O_5$, $M_w = 330.33$, monoclinic, space group $P2_1$, $a = 9.872(5)$ Å, $b = 4.921(3)$ Å, $c = 16.859(9)$ Å, $\beta = 96.518(9)°$, $V = 813.8(8)$ Å^3, $Z = 2$, $D_{calc} = 1.348$ g cm^{-3}. The crystal diffracted to a maximum angle of 23.6° in θ (a resolution limit of 0.91 Å). The data collection was performed using three different ω scans with exposure times of 50 s per frame and scan widths of 0.3° yielding data with an average completeness of 98.6% in the θ range 3.0 to 23.58°.

For both **1** and **2**, single crystal X-ray data were collected on a Bruker D8 diffractometer with an APEX CCD detector, and 1.5 kW graphite monochromated Mo radiation. The detector to crystal distance was 60 mm. The frames were integrated with the SAINT v6.45a.[31] A semi-empirical absorption correction using multiple-reflections was carried out using the program SADABS V2008-1.[32] The structure was solved and refined with X-SEED,[33] a graphical interface to SHELX97.[34] In the final cycles of refinement all non-hydrogen atoms were refined anisotropically, and the hydrogen atoms refined freely without restraints.

Powder X-ray diffraction

Powder X-ray diffraction data were collected on a Panalytical X'pert pro multi-purpose diffractometer (MPD) in reflection Bragg–Brentano geometry operating with a Cu anode at 40 kV and 40 mA. Samples were mounted as loose powder onto a silicon zero background holder. PXRD patterns were collected in 16 1 hour scans with a step size of 0.00657° 2θ and scan time of 115 s per step over 2–50° 2θ on a sample stage rotating at 2 s per rotation. The incident X-ray beam was conditioned with 0.04 rad Soller slits, automatic divergence slit (5 mm), mask (5 mm) and anti-scatter slit of 1°. The diffracted beam passed through an automatic antiscatter slit (5 mm), 0.04 rad Soller slits and Ni filter before processing by the PIXcel detector operating in scanning mode.

X-Ray fiber diffraction

Fiber diffraction samples were prepared *in situ* from dipeptide solutions with the addition of glucono-δ-lactone (GdL) for final pH targeting. By placing a 10 μL droplet between two wax filled capillary tubes and allowing the solution to gelate and air-dry, a partially aligned fibre sample was formed (as previously described[35]). The fiber sample was mounted on a goniometer head and X-ray diffraction data were collected using a Rigaku CuKα (wavelength 1.5419 Å) rotating anode and RAxis IV++ detector with exposure times of 20 minutes and specimen to detector distances of 200 mm. The diffraction images were examined using CLEARER[36] and the diffraction signal positions determined using a module within CLEARER. One-dimensional traces of signal positions and intensities were produced by defining the equator in CLEARER and sampling signal intensities within an angular search width of 60° as a function of distance from the centre of the pattern in pixels. These were then converted to *d*-spacing (Å) by Bragg's law. Simulated fiber diffraction patterns were calculated using CLEARER. Unit cell dimensions, space groups and atomic coordinates were used to generate a 'fiber' texture (cylindrically averaged around the fiber axis). Atomic coordinates were used to calculate the intensity and position of all reflections in reciprocal space for a true crystal and of these, which are intercepted by the Ewald sphere. Fiber axis direction, crystalline size, rotation and angular fiber disorder and diffraction geometry are then accounted for to trace diffraction intensities from each reciprocal-lattice point to the plane of the detector.[36] Values for fiber disorder were 0.4 rad and crystallite size 40 nm.

For the study of the hydration level of the fibers, fiber diffraction alignments were prepared as described and dehydrated by vacuum desiccation for a total of 13 days. Fibers were visually inspected for damage by using a cross-polarising microscope. Rehydration was achieved by submersion of a dehydrated fiber in water for a total of 2 minutes. XRFD was conducted as described with all exposure times at 20 minutes

and contrast ratios of patterns constant for reliable comparison.

Computational

Putative crystal structures of both molecules were generated using global lattice energy minimization. The approach used for conformationally flexible molecules involves four steps: determining a set of molecular conformations; generating loosely packed crystal structures with each molecular conformation;[37] lattice energy minimization of the trial crystal structures allowing the molecular conformation to adjust to crystal packing forces and a final higher quality energy calculation for each crystal structure.

The molecular geometries used to generate the initial trial crystal structures were based on a B3LYP/6-31G** optimized geometry for each molecule. The Mogul knowledge base of molecular geometries derived from the Cambridge Structural Database[38] was used to assess the most likely torsion angles around all acyclic bonds in each molecule and 144 starting conformations were chosen for each molecule, with torsion angles taken at the maxima in the torsion angle distributions found by Mogul. All 144 molecular geometries with which crystal structures were generated were derived from the same optimized molecular structure by rotation of the relevant torsion angles, with bond lengths and angles fixed.

Crystal structures were generated using a Monte Carlo simulated annealing search of the crystal energy landscape,[39] as implemented in the Accelrys Cerius2 software,[40] in which the molecular conformation is kept rigid. Parameters controlling the heating and cooling are those used in previous work.[41] Searches were performed in $P2_1$ and $P2_12_12_1$, the two most commonly observed space groups for enantiomerically pure molecular organic crystal structures. Initial lattice energy minimization of the resulting structures was performed allowing selected torsion angles to optimize, using the Dreiding force field.[42] Final energy calculations used an empirically parameterized *exp-6* intermolecular atom–atom model potential and a distributed multipole description of intermolecular electrostatics. Atom–atom parameters for C, N, O and non-polar H were taken from Williams *et al.*,[43,44] and parameters for polar hydrogen atoms, H(–N) and H(–O), were taken from Price and co-workers.[45,46] Atomic multipoles, up to hexadecapole on each atom, were calculated separately for the molecular conformation found in each crystal structure and were derived from a distributed multipole analysis of a B3LYP/6-31G** calculation. Polarization of the molecular charge density by its environment in the solid state was accounted for by performing this molecular calculation within a polarizable continuum, using the PCM model with a dielectric constant of $\varepsilon = 3.0$. The influence of this model for polarization on lattice energy calculations is described elsewhere.[47] The intramolecular energy of the final conformation in each crystal structure was also assessed from the same single point B3LYP/6-31G** PCM calculation. Around 5000 crystal structures of each dipeptide were energy minimized in this final stage of the calculations and the total crystal energy for each structure was calculated as the sum of the intermolecular energy and the relative intramolecular energy (taking the intramolecular energies relative to the most stable conformation found in any of the crystal structures). Duplicate crystal structures were located using the COMPACK algorithm,[48] comparing inter-atomic distances within clusters of 25 molecules taken from each crystal structure. From each set of duplicate structures, the one with the lowest total energy was kept and the higher energy duplicates were discarded.

Gelation studies

In all cases, a solution of the naphthalene-dipeptide was prepared by addition of **1** molar equivalent of NaOH (0.1 M) to a suspension of the dipeptide in water to give a final concentration of peptide of 0.5 wt%. A clear solution was formed after vigorous mixing for approximately 5 minutes. The pH was changed using glucono-δ-lactone (GdL)[20] as follows. GdL (6.4 mg) was added to a sample vial. An aliquot (0.5 mL) of the dipeptide solution was added and the solution agitated using a Stuart SA8 vortex mixer for 5 seconds. The sample was then allowed to stand undisturbed.

Results and discussion

A number of suitably functionalized dipeptides have been shown to be efficient hydrogelators. These include dipeptides functionalized at the *N*-terminus with a large, bulky aromatic group such as naphthalene[13,14,22] or fluorenylmethoxycarbonyl (Fmoc).[16–18,20,21] In our studies on a series of naphthalene-dipeptides,[22] we observed that very subtle changes in molecular structure had a profound effect on the outcome of the assembly process. Specifically here we discuss the assembly of two closely related naphthalene-dipeptides, **1** and **2** (Fig. 1). These peptides differ only in the order of the two amino acids, glycine–alanine for **1** and alanine–glycine for **2**. This results in the shift in

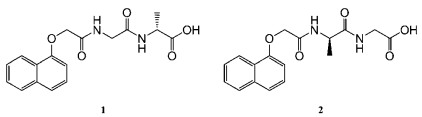


Fig. 1 Structures of functionalized dipeptides **1** and **2**.

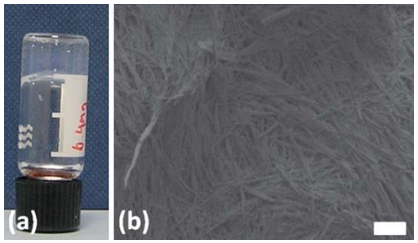


Fig. 2 (a) Photograph of self-supporting hydrogel formed from **1** on acidification. (b) The gel is formed *via* the entanglement of a fibrous network as demonstrated in the SEM image of the dried gel. The scale bar represents 2 μm.

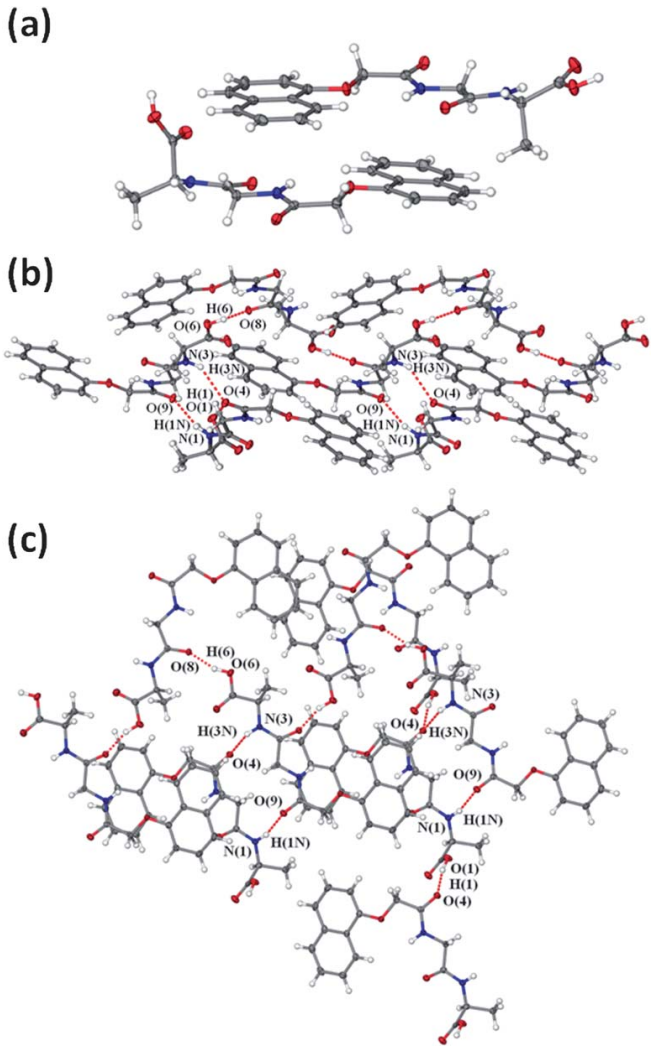


Fig. 3 Crystal structure of **1** crystallized from the gel phase. **1** crystallizes in the space group $P2_12_12_1$ with two molecules in the asymmetric unit (a). The two molecules differ in their conformation of the terminal carboxylic acid group. Molecules form an infinite 2D hydrogen bonded sheet in the crystal structure, viewed along 010 (b) and 001 (c). There are four intermolecular hydrogen bonds tethering the molecules in the *ab*-plane; these are indicated as dashed red lines.

a methyl group within the molecule. These peptides are related to those reported previously,[14] although the peptides here are coupled to the 1-position on the naphthalene ring rather than the 2-position.

Aqueous solutions of **1** and **2** were prepared *via* the addition of 1 molar equivalent of sodium hydroxide to a dispersion of the dipeptide in water. Assembly was then carried out by lowering the pH. The pH was lowered using glucono-δ-lactone (GdL) as described previously.[20–22] GdL hydrolyses slowly in water to give gluconic acid, resulting in a decrease in pH. Immediate differences were observed for solutions of **1** and **2**. On acidification, **1** initially formed a self-supporting hydrogel (Fig. 2). Crystallization from the gel phase occurred over 2 days, with some non-crystalline precipitate also forming. On the other hand, **2** always formed a crystalline precipitate on a decrease in pH. Again, GdL was used to decrease the pH in this system. We note that preparing crystals suitable for diffraction directly from solvents in which gels can be formed is extremely rare (most comparisons are between crystals grown in one solvent and gels formed in another)—this observation for **1** allows the comparison carried out here. The purpose of this work was to investigate the preferred solid state arrangements of these two molecules to help understand the observed differences in their behavior upon precipitation. The use of GdL to change the pH does not affect the bulk observations for both **1** and **2** compared to using HCl. Gels and crystalline materials similarly result in both cases; GdL has been used here as the slower pH changes result in crystals suitable for single crystal diffraction.

Crystal structures

Using GdL to carefully change the pH, crystals of **1** suitable for single crystal diffraction formed from the gel phase. To achieve this, multiple attempts were necessary, with 3.2 mg mL^{-1} of GdL found to be the most successful. Less GdL resulted in no crystallization with a mixture of precipitation and crystallization occurring from the gel phase at higher GdL concentrations. Interestingly, **1** crystallizes with two molecules in the asymmetric unit ($Z' = 2$); the two molecules lack symmetry due to different conformations of the terminal carboxylic acid group (Fig. 3a). Apart from the terminal carboxylic acid group, the molecule is remarkably planar, allowing stacking of the molecules. Although there are no face-to-face interactions between aromatic rings, the asymmetric unit consists of two nearly parallel molecules in a head-to-tail, tail-to-head arrangement forming a pair-wise face-to-face interaction (Fig. 3a) with a short separation between the two molecules (shortest distance 3.27(1) Å).

There are four intermolecular hydrogen bonds in the structure, with the following geometrical parameters (referring to atom labels in Fig. 3): O(1)–H(1)···O(4) (1.78(3) Å, 2.658(2) Å, 169(2)°); N(1)–H(1N)···O(9) (2.07(2) Å, 2.912(2) Å, 162.0(2)°); O(6)–H(6)···O(8) (1.74(3) Å, 2.643(2) Å, 160(3)°) and N(3)–H(3N) ···O(4) (2.36(2) Å, 3.119(2) Å,146(2)°). The difference in conformations of the two independent molecules is related to substantially different hydrogen bonding, with the carbonyl oxygen of an amide linkage accepting a hydrogen bond from an amide N–H group in one molecule, and the same carbonyl oxygen of the other molecule accepting a hydrogen bond from the carboxylic acid O–H group. The hydrogen bonds form an infinite 2-dimensional network (*i.e.* hydrogen bonded sheets) parallel to that along the *ab*-plane. A portion of this network is shown in Fig. 3. It is conjectured that the differences in environments and the complex network of intermolecular bonds are a result of packing molecules with significant differences in their local charge densities and therefore their potential intermolecular interactions—the naphthalene ring favors π-stacking and hydrophobic effects, while there are also strong hydrogen bond donors and acceptors.

Crystals of **2** suitable for single crystal diffraction were also grown. **2** is crystallized in the primitive monoclinic space group $P2_1$ ($Z' = 1$). Stacking of the naphthalene rings is evident (Fig. 4a), though there is a large offset (2.8 Å) between the rings along the stack such that there is only a small amount of the ring area available for π–π overlap. The distance between planes of the naphthalene rings is 3.235 Å, but the distance between centroids is 4.921 Å and is larger than expected for π–π stacking. Neighbouring columnar stacks of naphthalene rings interact *via* edge-to-face interactions, but these too are displaced with the hydrogen atom at an offset from the ring centroid, and with a relatively short C···H distance (2.663(10) Å). Four classic intermolecular hydrogen bonds were observed by examining the O and amide N environments. An intramolecular hydrogen bond does occur between the amide N of the alanine and the ether O atom on the naphthalene ring (N1–H1···O1, 2.01(6) Å, 2.596(9) Å, 116(4)°). The dipeptide molecules are linked by three intermolecular hydrogen bonds resulting in the formation of a 2-dimensional network, Fig. 4c. The intermolecular hydrogen bonding occurs between the H atoms of the terminal carboxylic acid groups and alanine carboxyl O atoms (O5–H5···O2, 1.95 Å, 2.775(7) Å, 167°), and between H atoms belonging to amide N atoms and amide group carbonyl O atoms (N2–H2N···O3, 2.01(4) Å, 2.973(10) Å, 167°).

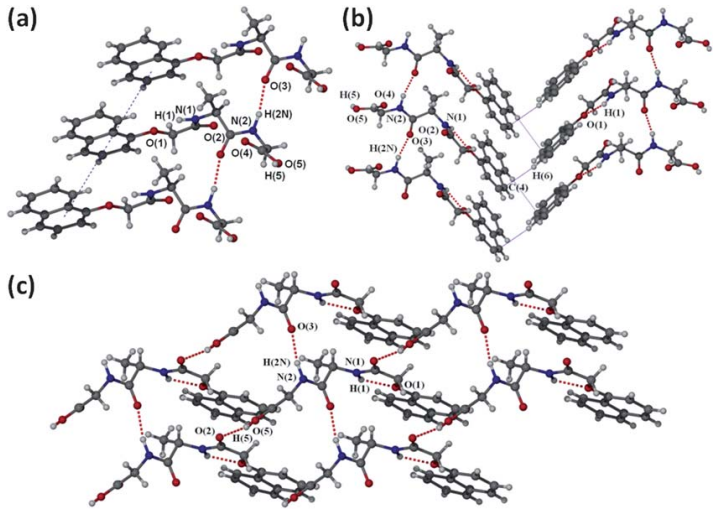


Fig. 4 Crystal structure of **2** crystallized from water. (a) Within a columnar stack, intermolecular hydrogen bonding is observed, and the naphthalene rings are offset with distances between the centroids greater than expected for efficient π–π interactions. (b) The structure is also stabilized by the displaced edge to face π···H interactions indicated by pink lines. (c) The intermolecular hydrogen-bonding (shown by dashed red lines) results in 2D sheets.

Gel structure

A number of techniques have been used to probe the molecular packing in the gel state, but results are complicated by the

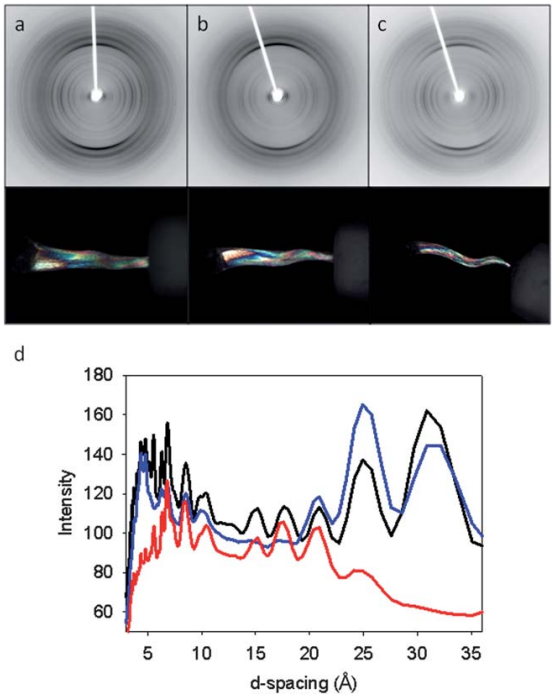


Fig. 5 X-Ray fiber diffraction from gels formed from dipeptide **1** after alignment (a), followed by dehydration by vacuum desiccation (b) and rehydration (c), fiber axes are vertical. Corresponding pictures show the birefringent alignments under cross-polarisation. The 1D trace shows the changes in equatorial reflections with de/rehydration (d). Black data correspond to (a), blue to (b) and red to (c).

hierarchical and complex structures formed. Whilst in the crystalline phase information is readily available from single crystal X-ray crystallography or powder X-ray diffraction (PXRD), the scattering contribution from the solvent molecules means that recording good quality diffraction data directly from a gel sample is often difficult. On the other hand, X-ray fiber diffraction (XRFD) can be collected from "*in situ*" prepared fiber alignments (prepared by drying a solution between the ends of two waxed capillaries).[35] Hence, for gels formed from **1**, a well-aligned fiber was prepared whose oriented diffraction (Fig. 5a) shows distinction between meridional and equatorial intensities.

The diffuse scattering observed in the XRFD pattern (Fig. 5a) suggests the fiber is hydrated, as would be expected for a hydrogel. Further, the hydration state of the fiber under quiescent conditions shows no change over extended periods of time as indicated by a comparison of XRFD patterns taken up to thirty-one weeks after gelation (data not shown). To elucidate whether the water is integral to the unit cell or part of a hydrated jacket surrounding the fibrils, XRFD was performed on an alignment following vacuum desiccation. Fig. 5a and b shows no appreciable change in morphology or birefringence (high degree of order) upon dehydration and a slight morphological change upon rehydration. Fig. 5a and b shows the subtle change observed in diffraction upon extensive dehydration. The major meridional 4.76 Å arising from a repetitive spacing along the fiber axis is unchanged. Equatorial reflections may be lost or relative intensities decrease to below detectable levels at 18, 15, 5.5 and 5.1 Å, but the majority of other reflections are unchanged in position. The significance of the reflections that are lost will require ongoing investigations. However, the preservation of the majority of reflections suggests that major structural features are preserved despite dehydration and thus although the hydration of the fiber is important, the structural features that give rise to the XRFD pattern do not involve water. Interestingly rehydrating the fiber (Fig. 5c) and examining by XRFD reveals the restoration of the lost reflections and a modulation of the relative intensity of reflections as shown by Fig. 5d. Damage to a unit cell dependent on the ordering of molecular water would likely result in a complete degradation of any structural architecture/order and thus loss of diffraction. When this is considered along with the ability to restore the original pattern with rehydration, the data are most consistent with a rearrangement of fiber packing. Further these data indicate that the fibers produced for XRFD are representatives of the structures in the native gel where they are fully hydrated. We have further attempted to determine the water content of the fiber from which the diffraction data were collected. The low mass of the fiber (approximately 0.2 mg) makes determination by TGA difficult.

The *d*-spacings from the fiber sample differ from those expected from the $Z' = 2$ single crystal structure (Fig. 6), suggesting that the underlying packing of the molecules in the two forms differs. We conclude that the molecular packing in the gel differs from that of the single crystal.

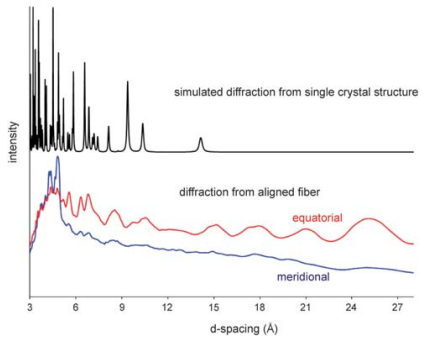


Fig. 6 Comparison of the simulated diffraction pattern from the crystal structure of dipeptide **1** (top) and measured diffraction from the aligned fiber (XRFD data shown in Fig. 5a).

Computational studies

A key question is why such different structures are formed from dipeptides **1** and **2**. Clearly, **1** and **2** have different molecular structures, but their close structural similarity means that they could adopt very similar conformations, both in the solution and the gel or crystal. It is also clear from above that **1** can aggregate to form different structures (bulk crystals, fibers) depending on the kinetic pathway. As noted in the Introduction, equilibrium kinetics predicts that the stable thermodynamic phase will be crystalline, with the gel phase representing a kinetically trapped phase. We have therefore employed global lattice energy minimization methods to predict the low energy crystal packing arrangements of both molecules. These calculations involve a sampling of possible crystal packing possibilities, with different molecular conformations, unit cells and space group symmetries. The aim is to generate an energy-ranked list of the most favorable crystal packing possibilities of a molecule—the crystal energy landscape. These methods are reliable at generating the observed crystal structure as one of a handful of the lowest energy calculated possibilities for small, rigid molecules[49] and recent progress has extended the applicability of these methods to more complex and flexible molecules.[47,50,51] Such calculations have previously been used to propose the structure of a xerogel formed from small rigid molecules.[52] This computational approach predicts the final crystal packing possibilities and does not take into account the kinetic pathway. However, the structures on these landscapes can hold information on likely intermolecular interactions that could dictate the kinetics. Therefore, we hypothesised that this approach would be instructive regarding the differences in observed behaviour between dipeptides **1** and **2**, *i.e.* why **1** forms a metastable gel, whilst **2** only forms a crystalline phase.

The conformational flexibility of the molecules studied here is at the limit of current capabilities of crystal structure prediction. For this reason, we were unable to study the further complexity introduced by the possible presence of water in the crystal or gel fibers, which could participate in or disrupt hydrogen bonding networks formed amongst dipeptide molecules. Using the current crystal structure prediction methodology, including water in the predictions would require a knowledge of water : dipeptide stoichiometry; hence, a measured water content would be

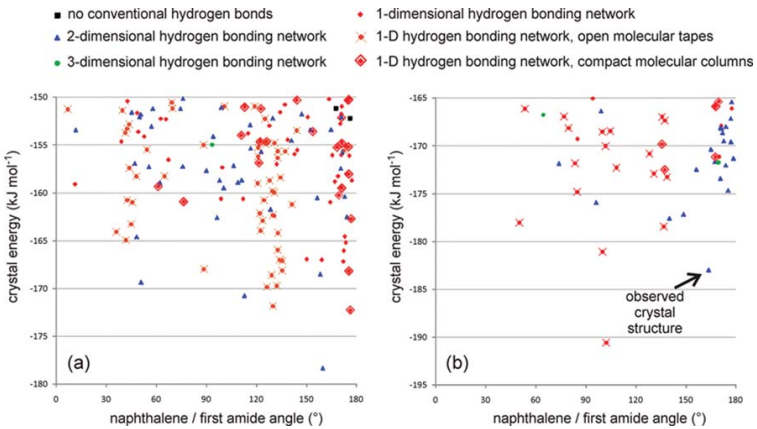


Fig. 7 Distributions of the low energy computer-generated crystal structures of (a) dipeptide **1** and (b) dipeptide **2**. Each point corresponds to a distinct lattice energy minimized crystal structure. The total crystal energies are plotted against the dihedral angle between the plane of the naphthalene and the HNCO plane of the first amide. The crystal structures are also classified by the dimensionality of the hydrogen bonding network.

required to inform the computational approach. The data shown above in Fig. 5 imply that water is not integral to the packing of the dipeptide-conjugates and is not intimately part of the structure. Therefore in the current computational approach, we assume that differences in conformational preferences and in dipeptide–dipeptide interactions are dominant in the assembly process.

The low energy regions of the predicted crystal energy landscapes of **1** and **2** are summarized in Fig. 7, where each data point corresponds to a distinct crystal structure. Such data are often presented as a plot of lattice energy against density. Here, the dihedral angle between the naphthalene and the plane of the first amide group is used as the horizontal axis in Fig. 7. This highlights the variation in molecular conformations among the predicted crystal structures, and the differences in conformational preferences between the two dipeptides. It has been proposed that this dihedral angle is important in determining the gel forming behavior of similar dipeptide molecules.[14]

The $Z' = 2$ crystal structure of dipeptide **1** could not have been found in the calculations performed here, which only considered crystal symmetries with one molecule in the asymmetric unit. Lattice energy minimization of the single crystal structure gives an energy of -191.2 kJ mol^{-1}: the $Z' = 2$ crystal structure is significantly lower in energy than any of the $Z' = 1$ predicted crystal structures (Fig. 7a), showing that there is a thermodynamic benefit for this molecule to crystallize with two independent molecules. However, as noted above, crystals of this form were very difficult to obtain, suggesting poor kinetics of nucleation or growth of this crystal form.

For dipeptide **2**, a very close match to the observed crystal structure was found in the list of predictions, as the second lowest energy structure. The unit cell lengths are all slightly overestimated, but agree with the observed dimensions to within about 2% (Table 1). The molecular conformation in the crystal structure and the arrangement of the molecules were predicted very closely (Fig. 8a and b), to a sufficient level that the structure could have been identified from the predictions on the basis of the similarity of its simulated PXRD pattern with the observed PXRD pattern of the bulk material (Fig. 8c). The presence of one lower energy predicted crystal structure of **2** might indicate that the observed packing is not the thermodynamically most stable possibility, and that the observed structure results from a kinetic preference during crystal nucleation or growth. However, we should be cautious in our conclusions; it is possible that errors in the relative calculated energies are as large as the energy differences between the predicted structures. Accurate calculated relative energies are a great challenge for crystal structures of highly flexible molecules, because of the difficulties in representing both intra- and intermolecular interactions in an accurate and balanced manner.[53]

The success of the predictions for molecule **2** validates the computational methods and provides confidence that the

Table 1 Unit cell dimensions of the observed and predicted crystal structures of dipeptide **2**. Both are in space group $P2_1$, with one molecule in the asymmetric unit ($Z' = 1$)

	a/Å	*b*/Å	*c*/Å	*b*/°	Unit cell volume/Å^3
X-Ray crystal structure ($T = 100$ K)	9.872(5)	4.921(3)	16.859(9)	96.518(9)	813.72
2nd lowest energy predicted structure	10.034(+1.6%)	5.029(+2.2%)	16.939(+0.5%)	98.36	845.69(+3.9%)

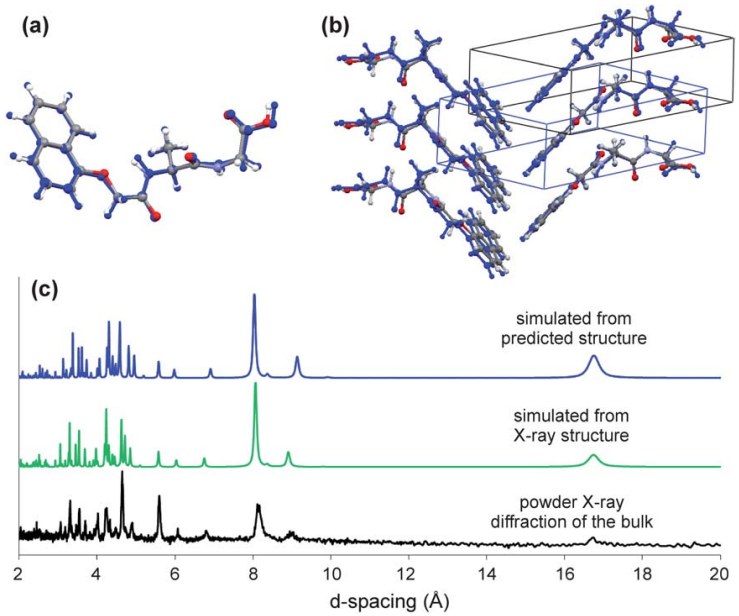


Fig. 8 Overlays of the X-ray determined and predicted structures of dipeptide **2**: (a) comparison of molecular conformations and (b) comparison of crystal packing. The X-ray structure is colored by element, and the black lines indicate its unit cell. The molecules and unit cell of the predicted structure are shown in dark blue. The offset in the unit cells is due to a different choice of unit cell origin in the observed and predicted structures. (c) Observed and simulated powder X-ray diffraction patterns.

predicted structures of the two dipeptides are representatives of the energetically favorable possibilities, so that similarities and differences between the two sets of structures provide a comparison of the packing preferences of the two molecules. Comparisons between the structure sets for **1** and **2** might shed light on the difference in gel-forming ability caused by such a small difference in molecular structure. The observed $Z' = 2$ crystal structure of **1** highlights the limitation that the crystal structure calculations were performed in a restricted region of crystal packing space and that other low energy possibilities will exist with different symmetries. We also note again that we have ignored the presence of water for the purposes of the calculation. Nonetheless, the success of the predictions demonstrates that the computational methods do provide physically realistic structures for these systems and that the chosen model of energies is faithful to the true interactions between dipeptides.

As a first classification of structures, all low energy predicted crystal structures were categorized according to the dimensionality of the hydrogen bonding network within the crystal (Fig. 7). Hydrogen bonds are the strongest and most directional intermolecular interactions in crystals of neutral organic molecules, and their arrangement has an important impact on the properties of the resulting material. Of specific interest here is the influence that the topology of strong interactions might have on the growth morphology, possibly leading to fiber growth and gelation. Hydrogen bonding can lead to fast crystal growth in a particular direction, so anisotropy of the strength and connectivity of hydrogen bonding in different directions through the structure can lead to anisotropy in crystal growth. The importance of one-dimensional assembly for gel formation was highlighted in a recent review;[26] hydrogen bonding that extends in only one direction through the crystal is most likely to lead to fast growth in only one direction and the growth of fibers.

While there is a two-dimensional hydrogen bond network in the global minimum energy predicted structure for dipeptide **1**, most of the other low energy structures for this molecule contain one-dimensional hydrogen bonding networks: 77% (50 of 65) of the predicted structures within 20 kJ mol^{-1} of the global energy minimum contain hydrogen bonds that extend in only one dimension through the crystal. The crystal packing landscape of **2** is much less dominated by structures with one-dimensional hydrogen bonding: such structures make up 52% (13 of 25) of the predicted structures within 20 kJ mol^{-1} of the global energy minimum. These differences in the distributions of hydrogen bond networks in the predicted structures already suggest that the small change in molecular structure between **1** and **2**

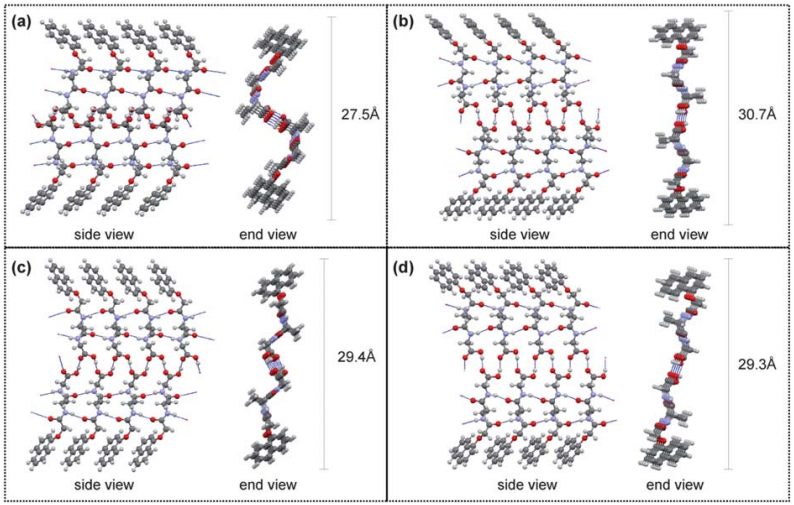


Fig. 9 Commonly observed 1-dimensional hydrogen bonded open molecular tapes in the low energy predicted crystal structures of dipeptide **1** (a and b) and dipeptide **2** (c and d). Hydrogen bonds are indicated by dashed blue lines.

significantly affects the preferred molecular packing and topology of intermolecular interactions.

We find that most of the predicted structures with one-dimensional hydrogen bonding networks can be grouped into two classes: structures with open tapes of hydrogen bonded molecules and structures with compact columns of molecules. Examples of open tapes found in the predicted crystal structures are shown in Fig. 9; this is the most common class of structure for both dipeptides. These structures involve pairs of amide–carbonyl hydrogen bonds between neighboring dipeptide molecules, resulting in chains that resemble a parallel β-sheet arrangement. These chains of molecules are associated *via* hydrogen bonding of the terminal carboxylic acid groups, extending the β-sheet structures into kinked molecular tapes with a typical width of around 30 Å (Fig. 9). This packing of molecules is further stabilized by π–π stacking of the naphthalene groups at both edges of the tapes. Of the structures within 20 kJ mol^{-1} of the global energy minimum, 56% (Fig. 7a) and 40% (Fig. 7b) of the predicted structures contain these hydrogen-bonded open tapes, for dipeptides **1** and **2** respectively.

The second commonly observed one-dimensional hydrogen bonding arrangement is a more compact packing, where the molecules aggregate into hydrogen bonded columns (two examples are shown in Fig. 10). As in the open tapes, these structures involve amide–carbonyl hydrogen bonds and π–π stacking of naphthalene rings. As in the tapes, some of these structures (*e.g.* Fig. 10b) involve β-sheet hydrogen bonding between parallel dipeptides, resulting in the typical β-sheet 4.8 Å repeat distance between dipeptides. However, the molecular conformation in these columns is usually less extended than in the open tapes. Hydrogen bonding between terminal acid groups is found in some of these structures, but the terminal acid group is more often hydrogen bonded to either the ether or first amide oxygen atom, so that the peptide chains overlap (Fig. 10). This overlap of the dipeptide chains, along with the folded molecular conformation, results in very compact columns of molecules, where the strong interactions (hydrogen bonding and π–π stacking) are aligned along the column. The hydrogen bonding in these predicted structures is similar to that proposed by Yang and co-workers based on transmission electron micrographs, circular dichroism and fluorescence spectra of the gels of very similar naphthalene dipeptides to those studied here.[14] The nature of the intermolecular interactions is also similar to those that are proposed to make up the fibers in Fmoc–diphenylalanine hydrogels,[17] except that their β-sheets involve anti-parallel aligned dipeptides, with Fmoc groups on alternating sides of the dipeptides, leading to an interlocking of the Fmoc groups between neighbouring β-sheets. The parallel alignment of dipeptides in our predicted structures allows for close π–π stacking within a single chain of molecules. This, combined with the smaller size of the dipeptide side groups here compared to the Fmoc–diphenylalanines, allows a more tightly coiled helix and overall smaller dimensions of the supramolecular assembly: cross-sectional dimensions of the hydrogen bonded columns found in our predictions are about 10–12 Å in the short dimension and 18–21 Å in the longer dimension (Fig. 10).

The alignment of all strong interactions (hydrogen bonding and π–π stacking) along these columns could promote fast

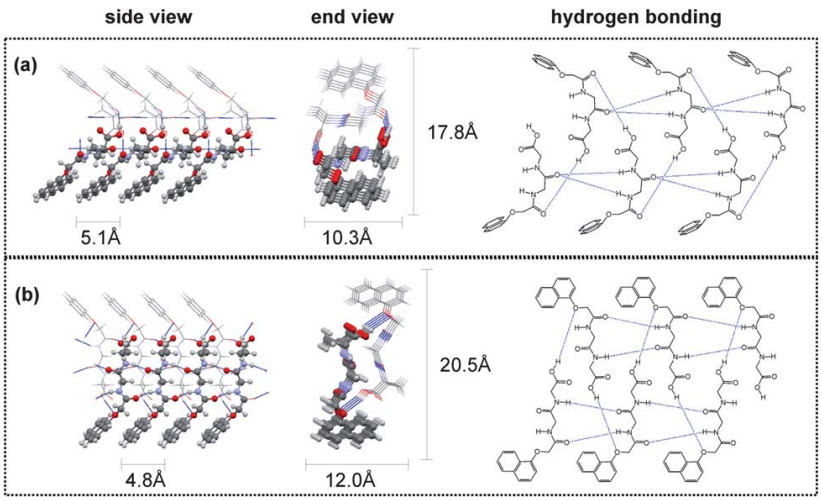


Fig. 10 Examples of the one-dimensional compact molecular columns found in the predicted crystal structures of dipeptide **1**. Hydrogen bonds are indicated as dashed blue lines. Diagrams on the left show the full 3-dimensional structure of the columns, while the right-hand diagrams show the hydrogen bond connectivity. The two structures shown here are found at −172.3 kJ mol^{-1} (a) and −156.9 kJ mol^{-1} (b).

aggregation in this direction, encouraging fiber growth and gelation. We find that these hydrogen bonded columns are more prevalent in the low energy region for dipeptide **1** than **2** (Fig. 7). Furthermore, these structures are closer to the global lattice energy minimum in the predictions for dipeptide **1** than for **2**. The best of these structures in the predictions of dipeptide **2** is 18 kJ mol^{-1} less stable than the lowest energy predicted structure. In comparison, for dipeptide **1**, the best of these compact structures is only 6 kJ mol^{-1} above the global minimum energy predicted structure. Overall, the calculations predict that aggregation into these compact columns is a preferred mode of packing for **1**, but is less favorable for **2**. If these hydrogen bonded columns are related to the molecular arrangement in the gel fibers, the predicted difference in likelihood of forming these columns could provide an explanation for the different behavior of the two molecules. We note that these structures have similarities to the crystal structure solved for dibenzoyl-L-cystine,[54] where the filament structure was related to gel forming ability. The dipeptides described here have pairs of hydrogen bonds running along single strands, which would be expected to strengthen the structures even further compared to the cystine example.

The strongest meridional reflection in the diffraction pattern from the gel fiber of dipeptide **1** occurs at about 4.76 Å (Fig. 6), corresponding closely to the typical β-sheet repeat distance between dipeptides and the repeat distance along the hydrogen bonded columns in the predicted structures (the repeat distance lies between 4.8 and 5.0 Å in the predicted structures). The meridional reflection would indicate that the columns are oriented along the axis of the fiber. The higher *d*-spacing reflections from the gel fiber are strongest in the equatorial direction and are related to repeat distances perpendicular to the fiber axis, possibly the relative arrangement of the hydrogen bonded columns. The relationship of the *d*-spacings of the equatorial reflections to the cross-sectional dimensions of the columns would depend on the arrangement of columns in a fiber. It is worth noting that two of those reflections are observed near 10 and 20 Å (Fig. 6), corresponding to the dimensions of these predicted hydrogen bonded columns (Fig. 10). The hydration state of the fiber may influence the packing of columns, with water most likely interacting with hydrogen bond donor or acceptor sites on the periphery of the columns. Intercalation of water between columns would be expected to have a strong influence on equatorial reflections above 10 Å, as we have observed.

It has been suggested that the angle between dipeptide backbone and aromatic ring is crucial in directing the assembly of gels in this class of molecules.[14] In the crystal structure of **1**, the torsion angle between the naphthalene ring and the first amide is 165° and 172° in the two independent molecules of the crystal structure of **1**, and is 169° in the crystal structure of **2**. In the crystal structures of both molecules, this angle is close to that expected to lead to an effective gelator.[14] The distributions of this angle in the low energy predicted structures show that both dipeptides can form low energy structures with a wide range of conformations (Fig. 7). However, there are a greater number of low energy structures with an angle near 180° for the gelating dipeptide **1**, and many of these structures are those with the compact one-dimensional hydrogen bonded columns described above. The conformational and hydrogen bonding preferences of the two molecules do seem to be correlated.

While it was not our intention to solve the structure of the fibers directly from the crystal structure prediction results, we did simulate powder X-ray diffraction patterns of all predicted structures of dipeptide **1** within 30 kJ mol^{-1} of the lowest energy prediction. These were compared to the observed fiber diffraction and fiber diffraction patterns were simulated from the predicted structures that looked most likely to match the observed diffraction from the fiber. None of the predicted structures gave a fiber diffraction that matched all of the observed reflections, indicating that the overall molecular arrangement in the gel is not identical to any of the predicted crystal structures. Nevertheless, we do observe similarities of the low *d*-spacing reflections in some predicted structures with those observed from the gel fiber. Hence, the predicted structures could be a starting point for a more complete structure solution. We are currently limited by the fact that predictions have been limited to $Z' = 1$ in common space groups. Clearly, the crystals that grow from the gel phase have $Z' = 2$ and the gel phase could also involve two or more dipeptide molecules in different conformations. The presence of water is another important consideration and this is currently being investigated further.

Conclusions

We have shown that two structurally similar naphthalene-dipeptides undergo self-assembly on acidification of an aqueous solution in different ways. One crystallizes on acidification whereas the other forms a hydrogel from which crystallization can occur over time. Slow pH adjustment with GdL results in crystallization from the gel phase over a number of hours. Fiber diffraction data of the gel phase demonstrate that a different packing of dipeptide molecules is present compared to the crystal phase.

Computational predictions of the crystal structures for these peptides were carried out to further understand these systems. The calculations accurately reproduce the crystal structure of the non-gelating dipeptide, though the observed $Z' = 2$ single crystal structure of the gelating molecule was outside the scope of the search. While we are unable to definitively determine the molecular arrangement in the gel fibers, the calculations do suggest important differences in conformational and hydrogen bonding preferences between the two molecules that could lead to their different aggregation behavior. The calculated structures for the gelating molecule also suggest packing that produces some of the observed diffraction features; the structures with these features are more favorable for **1** than **2**, compared to the rest of the predicted crystal packing possibilities. These overall differences in the low energy crystal structures of the two molecules provide some useful insight into how small changes in molecular structure can lead to dramatic changes in the behavior of molecules. We are still a long way from predicting from first principles whether a molecule will form a gel or crystallize. Nevertheless, we are encouraged by the results presented here that an analysis of the low energy crystal packing possibilities could be a helpful starting point for understanding the likely supramolecular arrangements and how these might influence molecular aggregation.

Acknowledgements

We thank the EPSRC (EP/G012741/1) for funding. GMD thanks the Royal Society for a University Research Fellowship.

References

1 M. de Loos, B. L. Feringa and J. H. van Esch, *Eur. J. Org. Chem.*, 2005, 3615–3631.
2 L. A. Estroff and A. D. Hamilton, *Chem. Rev.*, 2004, **104**, 1201–1217.
3 O. Gronwald and S. Shinkai, *Chem.–Eur. J.*, 2001, **7**, 4328–4334.
4 N. N. Adarsh, D. K. Kumar and P. Dastidar, *Tetrahedron*, 2007, **63**, 7386–7396.
5 M. J. Krysmann, V. Castelletto, A. Kelarakis, I. W. Hamley, R. A. Hule and D. J. Pochan, *Biochemistry*, 2008, **47**, 4597–4605.
6 R. P. Nagarkar, R. A. Hule, D. J. Pochan and J. P. Schneider, *J. Am. Chem. Soc.*, 2008, **130**, 4466–4474.
7 C. Veerman, K. Rajagopal, C. S. Palla, D. J. Pochan, J. P. Schneider and E. M. Furst, *Macromolecules*, 2006, **39**, 6608–6614.
8 A. Aggeli, M. Bell, N. Boden, L. M. Carrick and A. E. Strong, *Angew. Chem., Int. Ed.*, 2003, **42**, 5603–5606.
9 A. Aggeli, M. Bell, L. M. Carrick, C. W. G. Fishwick, R. Harding, P. J. Mawer, S. E. Radford, A. E. Strong and N. Boden, *J. Am. Chem. Soc.*, 2003, **125**, 9619–9628.
10 R. P. W. Davies, A. Aggeli, A. J. Beevers, N. Boden, L. M. Carrick, C. W. G. Fishwick, T. C. B. McLeish, I. Nyrkova and A. N. Semenov, *Supramol. Chem.*, 2006, **18**, 435–443.
11 A. Saiani, A. Mohammed, H. Frielinghaus, R. Collins, N. Hodson, C. M. Kielty, M. J. Sherratt and A. F. Miller, *Soft Matter*, 2009, **5**, 193–202.
12 E. F. Banwell, E. S. Abelardo, D. J. Adams, M. A. Birchall, A. Corrigan, A. M. Donald, M. Kirkland, L. C. Serpell, M. F. Butler and D. N. Woolfson, *Nat. Mater.*, 2009, **8**, 596–600.
13 G. L. Liang, Z. M. Yang, R. J. Zhang, L. H. Li, Y. J. Fan, Y. Kuang, Y. Gao, T. Wang, W. W. Lu and B. Xu, *Langmuir*, 2009, **25**, 8419–8422.
14 Z. M. Yang, G. L. Liang, M. L. Ma, Y. Gao and B. Xu, *J. Mater. Chem.*, 2007, **17**, 850–854.
15 Y. Zhang, H. W. Gu, Z. M. Yang and B. Xu, *J. Am. Chem. Soc.*, 2003, **125**, 13680–13681.
16 V. Jayawarna, M. Ali, T. A. Jowitt, A. E. Miller, A. Saiani, J. E. Gough and R. V. Ulijn, *Adv. Mater.*, 2006, **18**, 611–614.
17 A. M. Smith, R. J. Williams, C. Tang, P. Coppo, R. F. Collins, M. L. Turner, A. Saiani and R. V. Ulijn, *Adv. Mater.*, 2008, **20**, 37–41.
18 A. Mahler, M. Reches, M. Rechter, S. Cohen and E. Gazit, *Adv. Mater.*, 2006, **18**, 1365–1370.
19 R. Orbach, L. Adler-Abramovich, S. Zigerson, I. Mironi-Harpaz, D. Seliktar and E. Gazit, *Biomacromolecules*, 2009, **10**, 2646–2651.
20 D. J. Adams, M. F. Butler, W. F. Frith, M. Kirkland, L. Mullen and P. Sanderson, *Soft Matter*, 2009, **5**, 1856–1862.
21 D. J. Adams, L. M. Mullen, M. Berta, L. Chen and W. F. Frith, *Soft Matter*, 2010, **6**, 1971–1980.
22 L. Chen, K. Morris, A. Laybourn, D. Elias, M. R. Hicks, A. Rodger, L. C. Serpell and D. J. Adams, *Langmuir*, 2010, **26**, 5232–5242.
23 K. J. C. van Bommel, C. van der Pol, I. Muizebelt, A. Friggeri, A. Heeres, A. Meetsma, B. L. Feringa and J. van Esch, *Angew. Chem., Int. Ed.*, 2004, **43**, 1663–1667.
24 O. Gronwald, K. Sakurai, R. Luboradzki, T. Kimura and S. Shinkai, *Carbohydr. Res.*, 2001, **331**, 307–318.
25 F. M. Menger and K. L. Caran, *J. Am. Chem. Soc.*, 2000, **122**, 11679–11691.
26 P. Dastidar, *Chem. Soc. Rev.*, 2008, **37**, 2699–2715.
27 J. Cui, Z. Shen and X. Wan, *Langmuir*, 2010, **26**, 97–103.
28 Y. J. Wang, L. M. Tang and J. Yu, *Cryst. Growth Des.*, 2008, **8**, 884–889.
29 J. R. Moffat and D. K. Smith, *Chem. Commun.*, 2008, 2248–2250.
30 E. Ostuni, P. Kamaras and R. G. Weiss, *Angew. Chem., Int. Ed.*, 1996, **35**, 1324–1326.
31 *SAINT*, Bruker Inc., Madison, WI, USA, 2005.
32 *SADABS*, Bruker Inc., Madison, WI, USA, 2008.

View Online

Downloaded by University of Sussex on 17 February 2011
Published on 07 July 2010 on http://pubs.rsc.org | doi:10.1039/C0SM00409J

This journal is © The Royal Society of Chemistry 2010

33 L. J. Barbour, *J. Supramol. Chem.*, 2001, **1**, 189–191.
34 G. M. Sheldrick, *Acta Crystallogr., Sect. A: Found. Crystallogr.*, 2008, **64**, 112–122.
35 O. S. Makin and L. C. Serpell, in *Amyloid Proteins: Methods and Protocols*, ed. E. M. Sigurdsson, Humana Press, Totowa, 2005, pp. 67–80.
36 O. S. Makin, P. Sikorski and L. C. Serpell, *J. Appl. Cryst.*, 2007, **40**, 966–972.
37 G. M. Day, W. D. S. Motherwell and W. Jones, *Phys. Chem. Chem. Phys.*, 2007, **9**, 1693–1704.
38 I. J. Bruno, J. C. Cole, M. Kessler, J. Luo, W. D. S. Motherwell, L. H. Purkis, B. R. Smith, R. Taylor, R. I. Cooper, S. E. Harris and A. G. Orpen, *J. Chem. Inf. Comput. Sci.*, 2004, **44**, 2133–2144.
39 H. R. Karfunkel, F. J. J. Leusen and R. J. Gdanitz, *J. Comput. Aided Mol. Des.*, 1994, **1**, 177–175.
40 *Cerius2, Version 4.6*, Accelrys Inc., San Diego, USA, 2001.
41 G. M. Day, J. Chisholm, N. Shan, W. D. S. Motherwell and W. Jones, *Cryst. Growth Des.*, 2004, **4**, 1327–1340.
42 S. L. Mayo, B. D. Olafson and W. A. Goddard, *J. Phys. Chem.*, 1990, **94**, 8897–8909.
43 S. R. Cox, L. Y. Hsu and D. E. Williams, *Acta Crystallogr., Sect. A: Found. Crystallogr.*, 1981, **37**, 293–301.
44 D. E. Williams and S. R. Cox, *Acta Crystallogr., Sect. B: Struct. Sci.*, 1984, **40**, 404–417.
45 T. Beyer and S. L. Price, *J. Phys. Chem. B*, 2000, **104**, 2647–2655.
46 D. S. Coombes, S. L. Price, D. J. Willock and M. Leslie, *J. Phys. Chem.*, 1996, **100**, 7352–7360.
47 T. G. Cooper, K. E. Hejczyk, W. Jones and G. M. Day, *J. Chem. Theor. Comput.*, 2008, **4**, 1795–1805.
48 J. A. Chisholm and S. Motherwell, *J. Appl. Cryst.*, 2005, **38**, 228–231.
49 G. M. Day, T. G. Cooper, A. J. Cruz-Cabeza, K. E. Hejczyk, H. L. Ammon, S. X. M. Boerrigter, J. S. Tan, R. G. Della Valle, E. Venuti, J. Jose, S. R. Gadre, G. R. Desiraju, T. S. Thakur, B. P. van Eijck, J. C. Facelli, V. E. Bazterra, M. B. Ferraro, D. W. M. Hofmann, M. A. Neumann, F. J. J. Leusen, J. Kendrick, S. L. Price, A. J. Misquitta, P. G. Karamertzanis, G. W. A. Welch, H. A. Scheraga, Y. A. Arnautova, M. U. Schmidt, J. van de Streek, A. K. Wolf and B. Schweizer, *Acta Crystallogr., Sect. B: Struct. Sci.*, 2009, **65**, 107–125.
50 T. G. Cooper, W. Jones, W. D. S. Motherwell and G. M. Day, *CrystEngComm*, 2007, **9**, 595–602.
51 P. G. Karamertzanis, A. V. Kazantsev, N. Issa, G. W. A. Welch, C. S. Adjiman, C. C. Pantelides and S. L. Price, *J. Chem. Theor. Comput.*, 2009, **5**, 1432–1448.
52 K. M. Anderson, G. M. Day, M. J. Paterson, P. Byrne, N. Clarke and J. W. Steed, *Angew. Chem., Int. Ed.*, 2008, **47**, 1058–1062.
53 P. G. Karamertzanis, G. M. Day, G. W. A. Welch, J. Kendrick, F. J. J. Leusen, M. A. Neumann and S. L. Price, *J. Chem. Phys.*, 2008, **128**, 244708.
54 F. M. Menger, Y. Yamasaki, K. K. Catlin and T. Nishimi, *Angew. Chem., Int. Ed.*, 1995, **34**, 585–586.